173 MAY 3 1938

Vol. 22, No. 5 May, 1938

PSYCHOLOGICAL REVIEW PUBLICATIONS

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER

UNIVERSITY OF PENNSYLVANIA

CONTENTS

An Experimental Test of Freud's Doctrine of the Relation of Hedonic Tone to Memory Revival: A. A. SHARP 395

An Experiment on the Disinhibition of Voluntary Responses: W. S. HUNTER 419

Fatigue of the Vibratory Sense: C. H. WEDELL AND S. B. CUMMINGS, JR. 429

Whole and Four Part Learning Thirty Two Unit Spider Mazes: T. W. COOK 439

Studies in Thermal Sensitivity: 5. The Reactions of Untrained Subjects to Simultaneous Warm-Cold Stimulation: W. L. JENKINS 451

Note on the Reliability and the Validity of the Group Judgment: M. G. PRESTON 462

Effect of Visual After-Sensations Upon Brain Potential Patterning Under Different Degrees of Attention: L. E. TRAVIS AND M. E. HALL 472

The Effect of Subcutaneous Injections of Benzedrine Sulfate on the Activity of White Rats: L. V. SEARLE AND C. W. BROWN 480

A Mathematical Analysis of the Experiments in Extra-Sensory Perception: D. L. HERR 491

PUBLISHED MONTHLY, TWO VOLUMES PER YEAR
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND OHIO STATE UNIVERSITY, COLUMBUS, OHIO

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN A. MCGEOCH, *Editor*
Wesleyan University

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $6.00 (Foreign, $6.25). Single copies, 60c.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears monthly (since January, 1937), two volumes per year, each volume of six numbers containing about 625 pages.
Subscription: $14.00 ($7.00 per volume; Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 560 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

COMBINATION RATES

Review and Bulletin: $10.00 (Foreign, $10.50).
Review and J. Exp. (2 vols.): $17.00 (Foreign, $17.75).
Bulletin and J. Exp. (2 vols.): $18.00 (Foreign, $18.75).
Review, Bulletin, and J. Exp. (2 vols.): $22.00 (Foreign, $23.00).

Subscriptions, orders, and business communications should be sent to

THE PSYCHOLOGICAL REVIEW COMPANY

THE OHIO STATE UNIVERSITY, COLUMBUS, OHIO

Journal of
Experimental Psychology

VOL. 22, NO. 5 MAY, 1938

AN EXPERIMENTAL TEST OF FREUD'S DOCTRINE OF THE RELATION OF HEDONIC TONE TO MEMORY REVIVAL

BY AGNES ARMINDA SHARP

Instructor in Abnormal Psychology, Rush Medical College; Chief Psychologist, Psychiatric Institute, The Municipal Court of Chicago

PROBLEM

This study [1] was designed as an experimental test of the truth of Freud's doctrine of the relation of hedonic tone to memory revival. As a basis for a critical evaluation of any experimental test of the doctrine, an explicit statement of the various assumptions underlying the theory is a first essential. Freud was primarily a clinician and not a systematist, and hence we find no explicit statement of several features of his theory in his writings. A knowledge of these must be derived from the context, and their nature is to some extent a matter of personal interpretation. We shall first state our opinion of the various implied tenets of the doctrine.

1. Freud assumes (1) that unpleasant experiences are less likely to be revived than are pleasant and neutral experiences, and (2) that pleasant experiences are more likely to be revived than are unpleasant and neutral experiences.

There is no dispute about this feature of Freud's doctrine. The process of excluding unpleasant and painful experiences

[1] From the Psychological Laboratory of The University of Chicago. The writer wishes to acknowledge her indebtedness to Dr. Harvey A. Carr, Chairman of the Department of Psychology at The University of Chicago, under whose direction this study was made.

from memory is labelled a repression mechanism, and this mechanism is one of the basic concepts of his psychoanalytic theory. Freud also refers to the memory enhancement of pleasant experiences, and phantasy and daydreaming are cited as examples. His implied use of this process of memory enhancement which along with repression plays a decisive role in Freud's conception of the causation and treatment of the psychoneuroses and psychoses is the generally accepted one of striving toward gaining pleasure. He refers to it less frequently, and he has not given it a specific label as in the case of repression.

2. Freud implicitly assumes that hedonic tone is merely one of the factors influencing the likelihood of recall. The relative likelihood of recall of a series of items may also be a function of a number of other variable factors. Pleasant items will be more frequently recalled, and unpleasant items less frequently recalled, than are neutral items only when all other factors influencing recall are equal. No experimental setup can be regarded as a decisive test of the truth of Freud's doctrine unless these residual factors are adequately controlled.

3. In our opinion, Freud tacitly assumed that the differential effect of hedonic tone upon memory revival will be manifested only after some interval of time, and that this effect tends to increase with time. Freud did not assume that the memory revival of an unpleasant experience is immediately inhibited once and for all. Repression is actually an after-expulsion. Time is required for the mechanism of repression to be effective. In this connection we may note that many of the experiments that have been performed to test the truth of Freud's doctrine measured the relative revivability of pleasant and unpleasant items after an interval of only a day, and in these cases negative results were generally obtained. Such experiments cannot be regarded as an adequate test of Freud's theory. Any decisive experiment must employ tests for delayed recall.

We cite several considerations in support of our interpretation of this feature of Freud's theory. (1) Freud's doctrine was derived from his clinical experience of the repression in

adult life of unpleasant childhood experiences. He was here dealing with the inability to recall after very long intervals of time. (2) Repression, according to Freud, is primarily a defense mechanism. It is a means of getting rid of perseverating memories of an unacceptable and emotionally disturbing nature. If the unpleasant experiences did not at first exhibit a considerable amount of perseverative revival, there would be no need for repression. One might say that an unpleasant experience is one that naturally exhibits a considerable amount of initial memory revival, and one whose revival is mentally disturbing. Repression is thus a device to get rid of these disturbing memories. The theory thus assumes, in our opinion, that unpleasant experiences exhibit a considerable degree of initial revival, and that their repression is attained only after some interval of time. (3) Repression is a continuous process. An unpleasant memory is not permanently banished by a single mental fiat. The tendency to revival is always present, and this tendency must be continually inhibited. This feature of Freud's doctrine is quite evident in the following quotation:

"The process of repression is not to be regarded as something that takes place once and for all, the results of which are permanent, as when some living thing has been killed and from that time onward is dead; on the contrary, repression demands a constant expenditure of energy, and if this is discontinued the success of the repression will be jeopardized, so that a fresh act of repression would be necessary. We may imagine that what is repressed exercises a continuous straining in the direction of consciousness so that the balance has to be kept by means of a steady counter-pressure." [2]

4. Repression of the unpleasant and the enhancement of the pleasant are regarded as universal mechanisms, *i.e.*, exhibited by all individuals normal, psychoneurotic, and psychotic. The doctrine was derived from Freud's clinical study of psychoneurotic and psychotic subjects, but it was later extended to normal subjects as is illustrated by his book,

[2] Freud, S., *Collected Papers*, London: Hogarth Press, 1925, Vol. IV, pp. 89–90.

The Psychopathology of Everyday Life.[3] Freud makes no statements as to the relative degree to which these mechanisms are exhibited by the two classes of subjects. His followers state that psychoneurotic and psychotic subjects exhibit a greater amount of unsuccessful repression than do normal subjects. Freud has said "unsuccessful repression will have more claim on our interest than that of repression which is eventually successful; the latter will for the most part elude our study.[4]

This raises the question as to the type of subject that should be employed in order to make a decisive test of the truth of Freud's doctrine. It would seem that the use of psychoneurotic and mildly psychotic subjects would be the more appropriate, and perhaps the use of both normal and abnormal types of subject is advisable.

5. Obviously any test of the doctrine must use memory materials that conform to Freud's conception of the nature of pleasantness and unpleasantness. Freud has not explicitly defined these terms, and hence their meaning must be derived from a study of the context in which they are used. In our opinion, Freud's usage differs considerably from that of the conventional academic type of psychology which regards a pleasant or unpleasant experience as one which contains a specific *affective element*. Freud uses the terms more in their popular sense, and an experience is labelled as pleasant or unpleasant primarily on the basis of its dynamic and emotional characteristics. Repressed experiences are usually termed complexes, and it is generally admitted that this term has certain dynamic and emotional implications. We shall state briefly our conception of this feature of Freud's doctrine.

An unpleasant experience is one which we shall hereafter refer to as an *unacceptable* one. It is an experience that is so out of harmony with the individual's ideals, values and established modes of conduct that it is mentally and emotionally disturbing, and hence stimulates him to attempt to get rid of it. After one has had such an experience, there is but one

[3] Freud, S., *The Psychopathology of Everyday Life*, New York, The Macmillan Company, 1914, p. 338.

[4] Freud, S., *Collected Papers*, London: Hogarth Press, 1925, Vol. IV, p. 92.

way to get rid of its memory revival, viz., to banish it from mind and attempt to prevent its future recurrence. Repression is a defense mechanism—a device used to get rid of emotionally disturbing memories. In other words, one might very well define an unpleasant memory as one which we tend to repress or forget.

A pleasant experience, on the other hand, is an *acceptable* one. It is an emotional experience that is so in harmony with our sense of values and established beliefs and modes of conduct that it stimulates us to prolong and repeat it. It is an experience which tends to elicit a positive adaptive reaction. A pleasant memorial experience is thus one which we attempt to dwell upon and prolong, and which we desire to repeat. A pleasant memory, by definition, is one toward which we tend to react in such a manner as to enhance the likelihood of its future revival.

Historical

A considerable number of experiments have been performed to test the truth of Freud's doctrine. The results of some of these experiments are cited in support of Freud's doctrine, while many others find no differential effect in the recall of pleasant and unpleasant materials. As crucial tests of Freud's theory, most of these experiments are defective in that they do not conform to one or more of the above mentioned requirements.

1. All of these studies have employed normal subjects. While the use of such subjects probably cannot be regarded as a fatal defect, yet in our opinion psychoneurotic and mildly psychotic subjects would be more appropriate; and probably the use of both normal and abnormal subjects is advisable.

2. A large number of these experiments have tested recall only after an interval of a day, and they have generally failed to find any differential effect for pleasant and unpleasant materials. As we have indicated, such experiments cannot be regarded as crucial tests of the truth of the theory.

3. A few experiments have failed to equate the other factors influencing the relative likelihood of recall for the two

types of material. In this type of experiment, the subject is asked to recall as many pleasant items from his previous experiences as is possible in a given interval of time. The test is then repeated for unpleasant items. More pleasant items are usually recalled in this type of test, and this fact has been advanced in support of Freud's theory. The data are not decisive, however, for the differential results may be due to other factors than the hedonic character of the experiences. The relative amounts recalled may be due in whole or in part to such factors as the unequal number of the original experiences, their character, their degree of fixation, and the intervals elapsing between the recall and the original experience.

4. The major defect of most of these experiments is the fact that they did not use memory materials that conform to Freud's conception of the nature of pleasantness and unpleasantness. In one experiment, for example, the pleasant list of words to be memorized contained such items as smooth, fragrant, sweet, airy, clean, soft, kiss, reward; while the unpleasant list was composed of such items as dirty, chilly, rough, guilt, funeral, scorn, bitter.

Of the studies purporting to test Freud's hypothesis, two seem pertinent to this experiment. Koch [5] gave ten ten-minute quizzes over a period of twenty-three days to three equivalent classes in educational psychology. Each quiz was graded and returned to the writer at the next meeting of the class. The subjects were instructed to mark in class on a scale of 1–5 to indicate their reactions to the grade received in terms of the dimension 'very happy—happy—discouraged.' The papers were again collected by the instructor. After the tenth quiz no further examinations were held. Five weeks later the subjects unexpectedly were asked to recall all 10 grades. The 152 subjects were divided into two groups: those who had and those who had not kept a written record of their grades. Both groups recalled the '1' grades best. That is to say that the pleasant, in general, were correctly recalled

[5] Koch, H. L., The influence of some affective factors upon recall, *J. General Psychology*, 1930, 4, 171–190.

relatively more frequently than unpleasant ones. The group that kept records recalled the '5' grades second best, while the ones who had no written record recalled the '2' grade second best. That is, grades reported as pleasant or unpleasant tended to be better remembered than those reported as stimulating no affective reaction. Indifferent grades were remembered most poorly. Koch states that there seemed to be a slight relationship between the possession of such traits as are listed in the Woodworth Wells Personal Data Sheet and the tendency to remember relatively poorly unpleasant grades.

Does Koch's study satisfy the Freudian criterion of pleasantness and unpleasantness? Are these really poignant experiences to the average undergraduate? Did the student report his real attitude toward his grades or did he report what ought to be his attitude according to academic standards? If the student received a 'C' when he expected 'C' he would not be discouraged. If he received an 'A' when he expected a 'C' he might be indifferent or elated according to his sense of values. His attitude to a grade received will be a function of its relation to what was expected and desired.

Stagner [6] was interested in finding the relation of the feeling tone aroused by an experience to the retention of that experience. He felt that it is possible that the factors influencing emotional reaction are likewise those which determine retention, rather than that the feeling tone itself is a factor influencing memory. To investigate this phenomenon he felt that the narration of personal experiences by the subject was the best suited experimental setup.

One hundred and one subjects were given mimeographed sheets on which blank lines were provided for a brief description of the most pleasant and the most unpleasant experience of the preceding 15 days. Below this were provided 10 lines for as many items of detail and fact (such as odors, sounds, images, etc.) associated with the experiences as the subjects could note down. Three weeks later a typewritten copy of his description was handed to each subject on the same kind of

[6] Stagner, R., The redintegration of pleasant and unpleasant experiences, *Amer. J. Psychol.*, 1931, 43, 1020–8.

sheet; the list of associated items had been left off. The subjects were instructed to fill in as many as they could recall. Accuracy and completeness were taken as a measure of the redintegration of the experience. Scoring was done on the basis of the ideas rather than the words. Associations not given previously were thrown out. A greater percentage of the items associated with pleasant experiences than with unpleasant was recalled, the difference being 3.37 times the sigma of the difference. There is a reliable difference between the completeness of redintegration of pleasantness and unpleasantness after three weeks. An active process of forgetting is suggested by the fact that there is a reliable difference (2.99) between the relative recencies of the pleasant and unpleasant events given, and a considerable difference in the relative fluency of associations of details for the two events. In recalling unpleasant experiences the tendency to inhibit the revival of associated experiences may be regarded as repression.

Stagner's setup is in harmony with Freud's two criteria: (1) that unpleasant experiences tend to be repressed while pleasant experiences are enhanced in recall, and (2) that time is necessary for successful repression. He deals not with the revival of the experience, but rather with which of these two experiences when revived in memory will arouse the most associations. His assumption is, of course, that the more associations the more completely is the experience recalled in memory. While these are real experiences we do not know whether they are pleasant and unpleasant in Freud's sense.

Procedure

In our experiment, the memory materials consisted of three lists of 15 paired associates each. One list was composed of acceptable items, one of unacceptable items, while those of the third list were neutral or indifferent in hedonic tone. These materials are given in Table I. Three groups of psychoneurotic subjects, each consisting of 23 individuals, were employed. Each group memorized one of these lists to the point of one successful repetition, and later was subjected to three successive oral recall tests on 2, 9, and 16 days after original learning. After the last recall test, the subjects relearned the list to the criterion of one perfect reproduction. The experiment was then repeated upon three groups of normal subjects, each consisting of 22 individuals. For each of the six groups we obtained the following records—

learning scores, relearning scores, and the number and character of the items correctly recalled for each test.

The learning method was adopted in order to equate for the three conditions such factors as length of list, degree of learning, and length of time intervening between learning and recall. Psychoneurotic subjects were employed in order to make a crucial test of the theory, and the experiment was repeated upon normal subjects in order to discover whether the phenomena of repression and enhancement are alike or different in respect to character or amount for the two types of subject. Successive recall tests distributed over some interval of time were employed in the hope of discovering the approximate time of the onset of repression and enhancement, and possibly of testing the assumption that the amount of repression or enhancement increases with time. In the absence of knowledge, our selection of intervals was purely arbitrary, and it is frankly recognized that a greater number of tests distributed over a much longer interval of time would have been advisable. However our choice of the maximum interval was determined by the practical difficulty of keeping experimental control of such a large number of subjects for any longer period of time.

A distinctive feature of our experiment is the choice of memory materials. Since psychoneurotic subjects were to be used for a decisive test of the theory, we wished to use memory materials that were definitely known to be acceptable, unacceptable or neutral to these subjects. We had case histories and clinical notes taken in connection with the examination and treatment of 130 psychoneurotic and mildly psychotic patients in the Psychiatric Clinic of Rush Medical College. These records were analyzed with a view to finding a list of items that were unacceptable to all of these patients. On the basis of this analysis, we selected the 15 unacceptable items given in Table I. We chose these items on the basis of the repeated statements in the records that each of these patients had found these items to be distinctly disturbing and troublesome. The lists of acceptable and neutral items were selected in a similar manner. Of the 130 patients, we were able to make contact with and to induce 69 to serve as subjects in the experiment. The items selected were thus known to be acceptable, unacceptable, or neutral to all of the psychoneurotic subjects.

TABLE I

SHOWING THREE TYPES OF MATERIALS

(1) Acceptable Material	(2) Unacceptable Material	(3) Neutral Material
1. developing memory	1. failing always	1. carrying baskets
2. gaining affection	2. seeking attention	2. leaning down
3. securing justice	3. feeling inferior	3. arranging flowers
4. enjoying life	4. fearing poison	4. coming inside
5. growing attractive	5. losing out	5. blending colors
6. succeeding often	6. blaming me	6. scrambling over
7. liking folks	7. knowing jealousy	7. humming songs
8. thinking clearly	8. hating everybody	8. drawing pictures
9. acting generously	9. disliking folks	9. flying kites
10. becoming important	10. crying frequently	10. opening cans
11. feeling better	11. forgetting everything	11. going places
12. winning respect	12. going insane	12. making things
13. living safely	13. desiring babies	13. twirling around
14. finding contentment	14. wanting marriage	14. getting out
15. laughing happily	15. avoiding passion	15. pointing upward

In this type of experiment, it is evident that the subject is engaged in the task of memorizing and recalling words that symbolize certain acceptable or unacceptable experiences. Since these words as words have no hedonic tone, it is entirely possible for a subject to memorize and recall such a list as abstracted from their personal hedonic significance. Especially may this occur in those experiments which use a series of nouns as memory material, and it is quite obvious that such abstraction will necessarily vitiate an experiment as a decisive test of Freud's theory. Consequently we wished to use as memory material a series of verbal symbols that are likely to be memorized and recalled on the basis of their personal hedonic significance. After canvassing various possibilities, we adopted the method of denoting these acceptable and unacceptable items of experience by a unitary phrase of paired associates, each consisting of a verb as the stimulus term and a noun or qualifying word as the response term. This mode of arrangement has several advantages. 1. A unitary phrase is much more concrete and particular in its significance than is a single word. 2. Only familiar words were employed, and familiarity is conducive to the use of meaning. 3. Verbs were chosen as the stimulus terms, because these verbs imply a subject and thus emphasize the personal significance of these items. 4. The use of the paired associates also enabled us to equate the three lists in respect to the length, number of syllables, degree of familiarity, and apparent difficulty of the response terms.

The normal subjects were taken from the writer's psychology class of adult employed persons. These subjects used the same lists as did the psychoneurotic subjects. Since these individuals were used primarily as a control test, no attempt was made to discover whether the lists of items were also acceptable or unacceptable to them. Inasmuch as the hedonic significance of these items exhibited a high degree of universality for the psychoneurotic type of subject, perhaps one is justified in assuming that the items possess much the same significance for the normal subjects.

For the purpose of this experiment it is essential that the three groups of each type of subject be approximately equal in respect to memorizing ability. We relied upon the method of chance selection to attain this result. The subjects were assigned to each of the three groups on the basis of the order in which they were tested. The first group thus consisted of the 1st, 4th, 10th, etc., subjects tested.

As a check on the composition of the groups, we secured at the time of testing the following records for each subject: chronological age (Otis Higher Form), I. Q. score, and a score on the Thurstone Personality Schedule. Table II gives the average values and standard deviations for each of the six groups and for the two types of subject. We also secured the medical diagnosis of each of the psychoneurotic or mildly psychotic individuals, and the composition of the three groups in respect to these diagnoses is given in Table III.

We may first note the differences between the *two classes* of subject. (1) The normal group has much the smaller personality score. This statement is true when we compare the averages for the two groups, and the score for each of the normal groups with the score for the corresponding abnormal group. This difference was to be expected, and it supports the diagnostic validity of this test. (2) The normal subjects have the larger I. Q. score in all four comparisons, and some of these differences exceed the limits of chance. (3) The normal subjects are the younger in three of the four comparisons. The exception refers to the two groups that learned the neutral materials. It may be noted that the difference between the group averages for each of the three categories exceeds the limits of chance. For the purposes of our experiment, however, it is not essential that the two groups be equal in respect to age, I. Q., personality score, or even learning ability. In fact, the close approximation of the

TABLE II

A COMPARISON OF THE NORMAL WITH THE ABNORMAL SUBJECTS

Subjects Normal

Material	N	Age	S.D.	I.Q.	S.D.	Personality	S.D.
1. Acceptable	22	26.6	5.97	104.	5.8	27.89	10.3
2. Neutral	22	29.54	9.1	110.	10.35	20.68	9.95
3. Unacceptable	22	24.72	5.1	108.5	9.15	16.94	7.6
Total	66						
Averages		26.95	7.18	107.5	5.94	21.83	10.14

Subjects Abnormal

Material	N	Age	S.D.	I.Q.	S.D.	Personality	S.D.
1. Acceptable	23	31.6	7.2	102.	9.0	89.6	16.08
2. Neutral	23	27.5	7.1	101.6	9.75	88.68	18.7
3. Unacceptable	23	31.69	5.56	104.7	7.65	75.9	11.0
Total	69						
Averages		30.26	6.94	102.4	8.8	84.72	15.48

TABLE III

SHOWING THE MEDICAL DIAGNOSIS OF THE ABNORMAL (EXPERIMENTAL) GROUPS

Diagnosis	Neutral Material: Number	Acceptable Material: Number	Unacceptable Material: Number
Alcoholic depression	1	1	1
Alcoholism and anxiety		1	1
Anxiety psychoneurosis	1	1	1
Compulsion neurosis	1	1	
Drug addiction	1	1	1
Fear of body odor			1
Functional colitis	1	1	1
Functional heart	1		1
Functional thyroid	1		1
Hysterical amnesia	1	1	
Hysterical fugue	1	1	1
Inferiority feelings	1	1	1
Manic-depressive	1	1	1
Dementia præcox (beginning)	1	1	1
Paranoid	3	2	3
Pathological lying	1	1	
Psychopathic personality		2	2
Persecution ideas		1	1
Psychopathia sexualis	1	1	
Schizoid personality		1	1
Schizophrenia (mild)	1	1	1
Situational depression	1	1	1
Stuttering	2		1
Suicide attempts	1	1	1
Functional torticollis	1	1	
Transportation phobia			1

two groups in respect to age, I. Q., and learning scores (Table IV) is quite surprising in view of the fact that they were drawn from different populations and that no attempt was made to equate them.

The attempt was made to equate the three groups of each class of subject by the method of chance selection, and hence we should expect that none of these differences will exceed the limits of chance. The data justify this expectation. All of the differences between the three psychoneurotic groups fall within the limits of chance variation, and the same statement is true for the three normal groups with the single exception of one of the differences in personality score. The data thus indicate that the three groups are also equated in respect to *memorizing ability* within the limits of chance variation.

One might have selected the three groups of each class of subject so as to attain an exact equation in respect to average age and I. Q. This method was not adopted for two reasons. (1) Age and I. Q. are independent variables, and it would have been practically impossible to equate the three groups in respect to both categories. (2) The essential requirement of the experiment is an approximate equation of the three groups in respect to *memorizing ability*. Hence an equation in respect to either age or I. Q. would attain this result only on the assumption of an approximately perfect correlation between one of these factors and memorizing ability. There is no evidence that these correlations are large enough so that an exact equation in respect to either age or I. Q. would give a better equation in respect to memorizing ability than the method of chance selection which we adopted.

All subjects were tested individually. We followed the usual procedure of the paired associates method. The 15 pairs of the list were serially presented, and each presentation was followed by a recall test. These alternate presentations and recalls were continued until the subject mastered the list up to the point of one perfect recall. For presentation, each pair of words was typed upon a separate white card. These cards were assembled in a pack, and the cards were manipulated by hand in a brass exposure box. The order in which the pairs were presented was kept constant from trial to trial. The duration of exposure of each pair was approximately 2 seconds, and the subject was required to pronounce both words during the period of exposure.

For the recall tests, we used a similar stack of cards each containing one of the stimulus terms. The order of presentation was varied from test to test by reshuffling the cards. The duration of exposure was 2 seconds, and the subject was requested to pronounce both the stimulus and the absent response term.

Results

The numerical results obtained are given in Table IV. This gives the mean score and its P.E. for each of the six groups for the number of trials to learn, the number of words recalled on each of the three successive recall tests, and the number of trials required to relearn the lists. The mean scores and their P.E.s are listed for each of the two classes of subjects.

Learning Scores.—The mean learning score of the normal subjects is slightly smaller than that of the psychoneurotic

subjects. Likewise the score for each of the three normal groups is smaller than that for the corresponding abnormal group. These differences cannot be functions of the character of the lists, for the lists were the same in all four comparisons. The normal subjects are thus slightly superior in respect to memorizing ability. While an exact equation of the two types of subject is not essential to the purpose of the experiment, we may note that they are much more alike in respect to memorizing ability than one would expect in view of the fact that they were drawn from different populations and no attempt was made to equate them.

TABLE IV

MEAN SCORES AND P.E. FOR LEARNING AND RECALL FOR SIX GROUPS OF SUBJECTS AND FOR TWO CLASSES OF SUBJECTS

Subjects	N	Material	Number of Trials to Learn	Number of Syllables 1st Recall	Number of Syllables 2nd Recall	Number of Syllables 3rd Recall	Number of Trials to Relearn
Normal.....	22	Acceptable	9.68±.630	8.68±.367	10.54±.530	10.18±.513	3.0 ±.294
Normal.....	22	Neutral	4.22±.190	10.04±.308	9.68±.378	9.13±.408	4.36±.342
Normal.....	22	Unacceptable	8.77±.422	6.45±.390	5.95±.396	5.72±.330	6.90±.406
Abnormal....	23	Acceptable	9.86±.819	9.17±.338	9.60±.403	9.17±.381	3.56±.256
Abnormal....	23	Neutral	6.52±.650	9.56±.276	8.56±.256	7.69±.361	4.39±.246
Abnormal....	23	Unacceptable	9.43±.596	5.86±.274	6.04±.492	5.86±.421	7.0 ±.550
Normal.....	66	3 Materials	7.57±.351	8.71±.239	8.7 ±.290	8.41±.082	4.75±.059
Abnormal....	69	3 Materials	8.60±.418	8.20±.220	7.9 ±.259	7.53±.067	4.95±.046

The three groups of each type of subject differ materially in respect to learning scores. In these comparisons, the differences may be functions of the groups or of the lists, and there is good evidence that both factors are operative. (1) The method of chance selection does not give an exact equation of the three groups; it equates them only within a certain range of chance error. The obtained differences in learning score must thus be accounted for in part in terms of group differences. (2) The character of the lists has also influenced the learning scores. This fact is attested by several lines of evidence. Some of the differences materially exceed the limits of chance variation. The differences between the scores for the neutral and the hedonic lists far exceed those obtained by comparing the two classes of subject for the same lists, and in the latter comparison no attempt was made to equate the two groups. The learning scores are not correlated with the

age and I. Q. records. Of the three normal groups, the one that learned the neutral lists was the oldest, had the highest I. Q., and made the lowest learning score. However, the abnormal group that learned the neutral list also made the lowest learning score, but its age and I. Q. scores were the smallest of the three groups. Finally the order of difficulty of the lists was exactly the same for the two classes of subject. In both cases, the neutral material was much the easiest to learn, and the acceptable material was slightly harder to learn than the unacceptable list. That is to say, the learning scores are correlated with the hedonic character of the lists and not with the group.

We may also note that the chance factor may operate to increase or decrease the size of the differences due to the hedonic character of the lists. Consequently in some of these comparisons, the list factor may be more efficacious than the obtained learning scores would indicate.

Recall Scores.—The recall scores were changed into percentage scores, and these values are graphically represented in Figs. 1 and 2. The continuous line gives the data for the unacceptable material, the dotted line for the neutral material, and the dash lines for the acceptable material. The initial value of 100 percent denotes the successful recall attained in the last learning trial.

These curves are like the conventional retention curve in that they exhibit an extremely rapid rate of initial forgetting. They exhibit a slower rate of forgetting for the two longer intervals than does the conventional curve. Any such difference is probably a function of the three successive recalls. Previous experiments have indicated that partial recalls tend to inhibit the subsequent rate of forgetting, and hence the values for the last two recalls were probably increased by the preceding recalls. However, the form of these curves as related to a conventional norm is not significant for our purposes, for we are interested merely in the differences between these six curves, and all were equally subject to the influence of this factor of successive recalls.

There are two outstanding characteristics of these recall scores. (1) The results are essentially identical for the normal and the psychoneurotic subjects, and (2) the amount recalled varies markedly with the list.

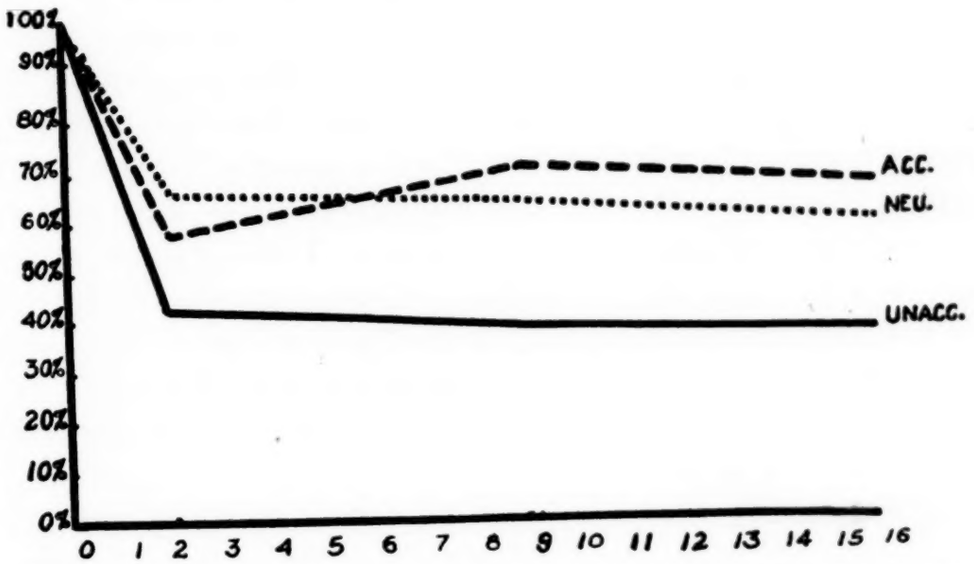


FIG. 1. Normal subjects. Percentages of recall for the three types of material.

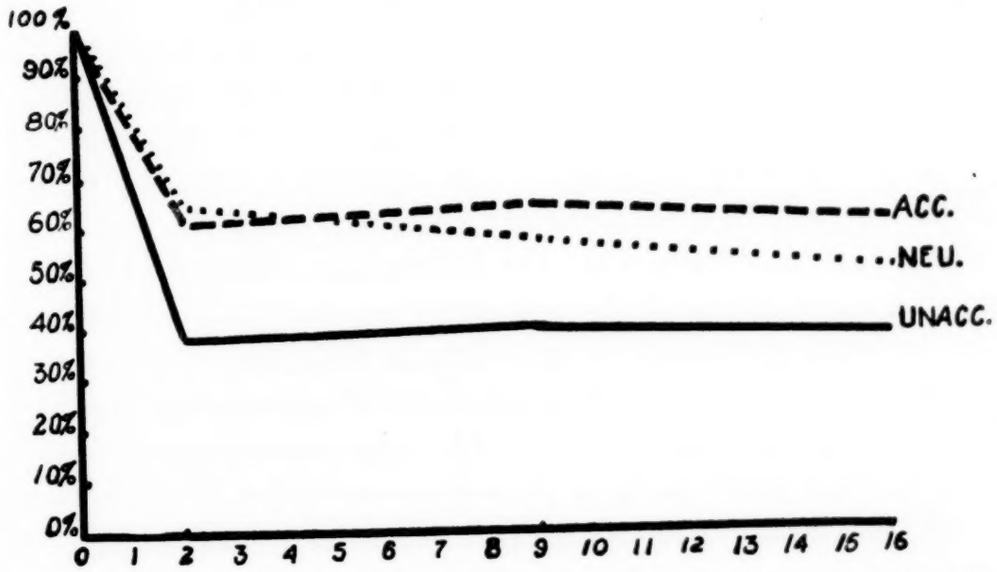


FIG. 2. Psychoneurotic subjects. Percentages of recall for the three types of material.

Both the normal and the psychoneurotic groups exhibit a large amount of repression of the unacceptable material relative to the neutral material. This repression is present at the two-day interval. Presumably the repressive mechanism

must begin to operate shortly after learning. There is some indication that the relative amount of repression may decrease for the psychoneurotic group after the two-day interval, but more tests distributed over longer intervals of time are essential to any final statement as to the existence of such a trend.

Both the normal and the psychoneurotic groups exhibit a small amount of repression of the acceptable material at the end of two days. The retention curves then rise, and both groups exhibit the phenomenon of enhancement for the 9- and 16-day intervals. The acceptable material is repressed at first and then enhanced in both groups. There is some indication that the relative amount of enhancement may increase for the psychoneurotic group with the passage of time, but any confident statement of the existence of such a trend will require a larger number of recall tests distributed over a much longer interval of time.

Theoretically the differences between the three retention curves for each type of subject may be functions of the groups or of the lists. Two facts indicate that they are primarily functions of the hedonic character of the lists. (1) The differences far exceed what may be reasonably expected on the basis of chance selection. (2) The same order of difference obtains for both the psychoneurotic and the normal subjects. In other words the differences are correlated with the nature of the lists and not with the groups.

We had also found that the differences between the learning scores are primarily functions of the character of the lists. The hedonic character of the lists is thus the prime determinant of both the learning and the recall scores, but it affects these two scores differently.[7] The unacceptable material is much more difficult to learn than the neutral list, but it is more poorly retained. The acceptable material is also more difficult than the neutral list, but it is more effectively retained for the longer intervals. The acceptable and unacceptable materials are approximately equal in difficulty, but they differ radically in respect to retention.

[7] As stated in the early description of this study there is a possibility of some other difference, other than hedonic tone, that was not equated and eliminated.

The differential effect of the hedonic character of the list upon the learning and recall scores is not inconsistent with what we know about the functional relation obtaining between these scores. It has long been known that these scores are related, but this relation is quite complex and it has as yet been inadequately analyzed. We shall mention briefly three principles involved in this relationship for which there is some experimental evidence. (1) When a given material is learned to the same criterion of mastery by individuals differing in ability, there is some evidence that the more proficient learners have the larger recall scores. Under these conditions, the learning and the recall scores tend to exhibit a negative correlation. (2) When individuals who are alike in ability learn the same materials by two different methods, *e.g.*, whole vs. part methods, massed vs. distributed practice, there is good evidence that the more efficient method gives the greater retention. This condition also gives a negative correlation between learning and recall scores. (3) When individuals who are alike in ability learn materials that are alike except in respect to length of list, there is good evidence that the learning and recall scores are positively correlated. Two of these conditions thus give a negative correlation, while the third mediates a positive correlation, and this analysis probably does not exhaust the complexities of this relationship. Hence any type of correlation may reasonably be expected for the conditions obtaining in our experiment.

Qualitative Differences in Recall.—The recall scores indicate that the two classes of subject do not differ materially in respect to the amount of repression and enhancement. However, it is possible that they do differ in respect to the items that were repressed or enhanced, and our data were further analyzed in reference to this possibility.

The repression and enhancement data for the 9-day interval are given in Figs. 3, 4, and 5. The items are distributed along the abscissa in order of their presentation during learning. The height of the graphs for each of the items represents the number of individuals in the group who recalled that item in the test. The continuous line gives the data for the psy-

choneurotic subjects, and the broken line for the normal subjects.

FIG. 3. Acceptable material. Individual successes in the second recall test of items by the two types of subjects.

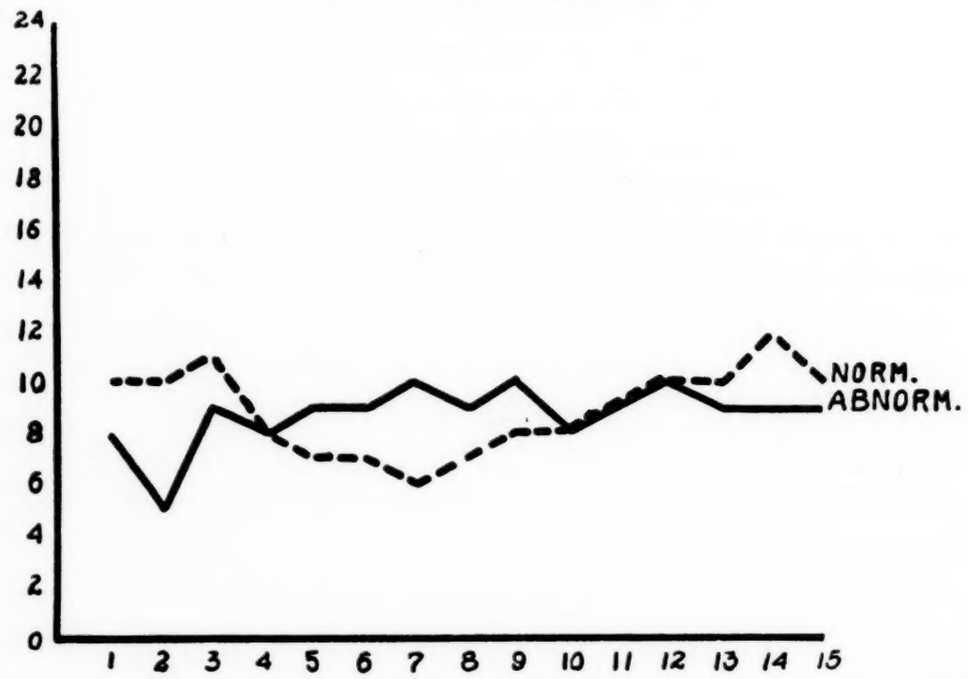


FIG. 4. Unacceptable material. Individual successes in the second recall test of items by the two types of subjects.

The curves give no decisive evidence that the two classes of subject differ in any significant fashion. We may thus

conclude that the two classes of subject do not differ significantly in respect to the character of the items that are repressed or enhanced.

We may note that these graphical data reveal one striking fact of some general significance although it has no bearing upon our experimental problem. For the acceptable and neutral materials, the relative likelihood of recalling any item is a function of its position in the list. The first and last items are most readily recalled, while those occupying the middle positions in the lists exhibit the least likelihood of

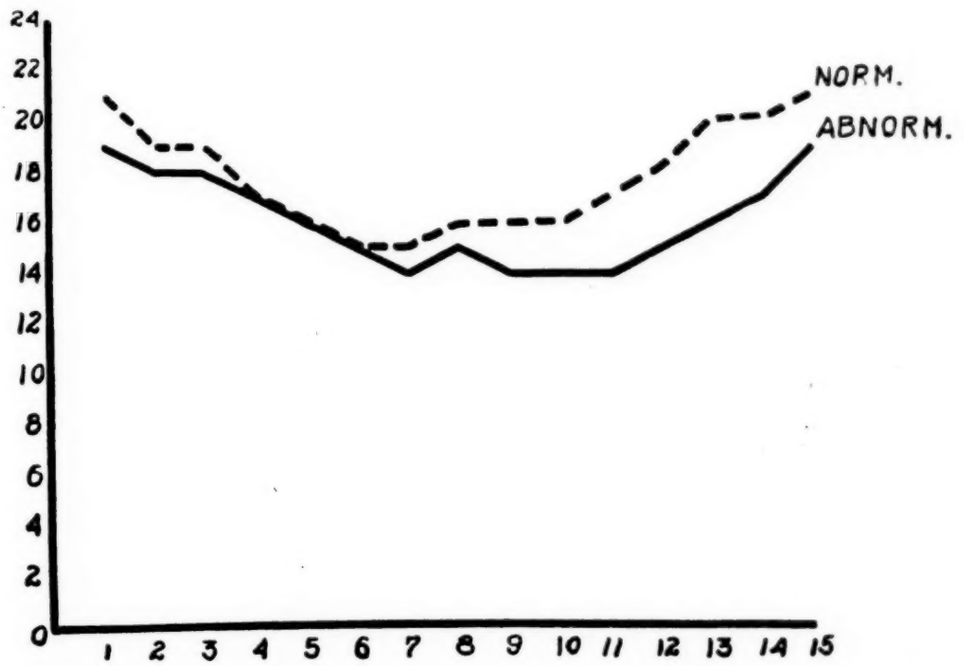


FIG. 5. Neutral material. Individual successes in the second recall test of items by the two types of subjects.

recall. The acceptable material exhibits this phenomenon to a greater degree than does the neutral list. The curves for the unacceptable material, however, give no decisive evidence that the relative likelihood of recalling any item varies with its position in the list.

Previous experiments on the so-called Laws of Primacy and Recency have shown that the rate of memorizing items is also a function of their position in the list. Consequently there is the possibility that likelihood of recall and rate of learning are correlated phenomena. To test this hypothesis, we computed the average number of trials required to memo-

rize each item. These data are graphically represented in Figs. 6, 7, and 8. The items are distributed along the abscissa in the order of their presentation. The ordinal values

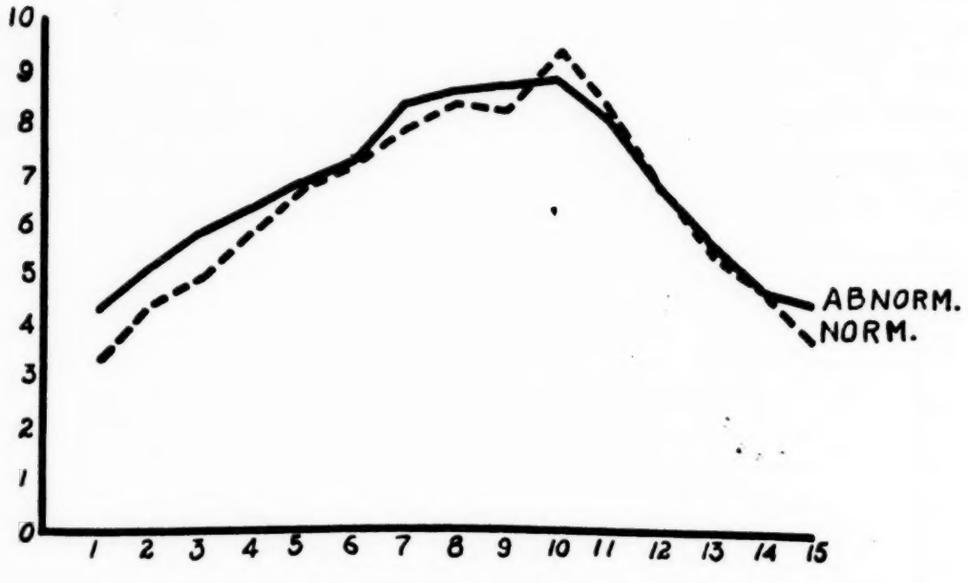


FIG. 6. Acceptable material. Average trials to learn each item by two types of subjects.

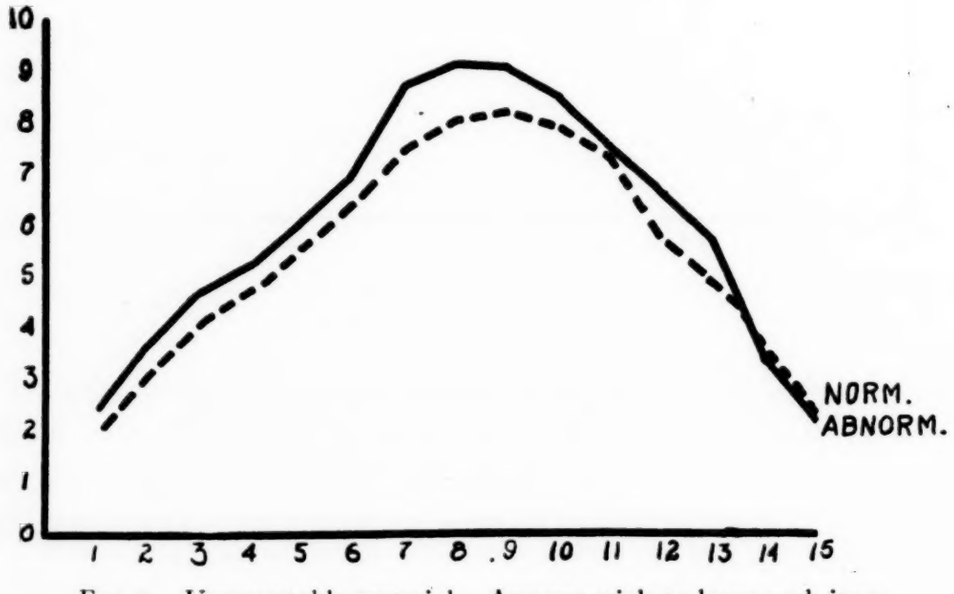


FIG. 7. Unacceptable material. Average trials to learn each item by two types of subjects.

denote the average number of trials required to learn each item. The continuous line gives the data for the psychoneurotic group, and the broken line for the normal subjects.

All six graphs exhibit the Primacy and Recency phenomena. The curves for the neutral material are relatively quite flat, but this fact is merely a function of the small number of trials required to memorize this list. Obviously the more difficult the list, the greater must be the maximal height of these curves.

A comparison of the two sets of graphs reveals the fact that relative likelihood of recall is correlated with relative rate of learning for the acceptable and neutral lists. No such relation obtains for the unacceptable material.

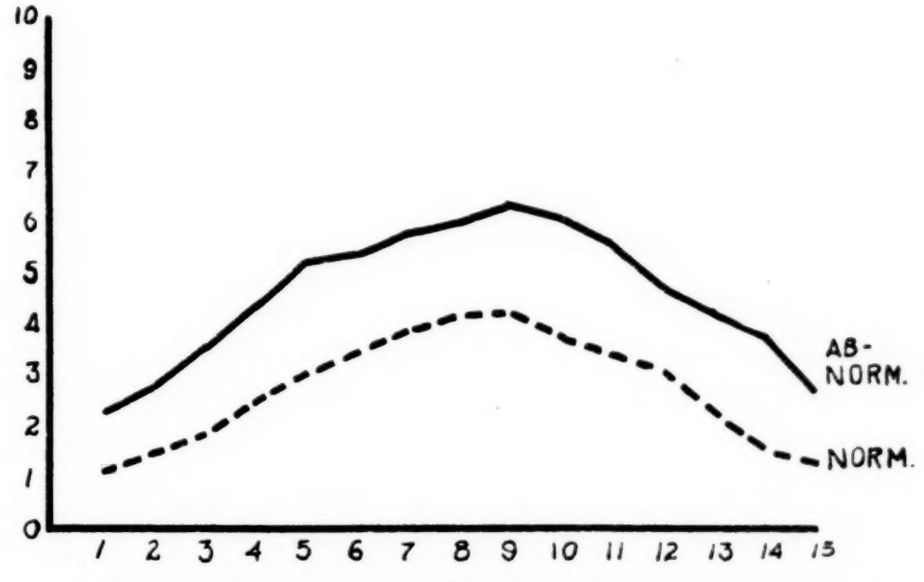


FIG. 8. Neutral material. Average trials to learn each item by two types of subjects.

There are two possible interpretations of this correlation. (1) We may assume that ease of recall is a function of ease of mastery. (2) All items continue to be presented and recalled after they are learned. Consequently we may assume that the items that are first learned exhibit the greatest degree of overlearning, and that ease of recall is a function of the degree of mastery. No reason can be assigned for the absence of this correlation for the unacceptable material.

Relearning Scores.—The relearning scores as given in Table II were transformed into saving scores by the usual formula. These saving scores purport to measure the amount retained in terms of the percentage of trials saved in relearning, while

the recall scores measure the amount retained in terms of the percentage of the total number of items that are recalled. We thus have the percentages of retention for the 16-day interval as measured by two different methods, and the comparative data are given in Table V.

TABLE V

SHOWING THE PERCENTAGE OF MATERIAL RETAINED AS MEASURED BY THE SAVING METHOD AND THE METHOD OF RECALL

Subjects	Material	% Recalled on 3rd Recall	% Retained—Saving Scores
Normal	Acceptable	67.8%	69%
Normal	Neutral	60.8%	−.003%
Normal	Unacceptable	38. %	21%
Abnormal	Acceptable	61. %	64%
Abnormal	Neutral	51.1%	32%
Abnormal	Unacceptable	39. %	25%

While the two methods purport to measure the same thing in comparable terms, yet it has long been known that they do not give identical values. For example, Luh found that the relearning method gave the smaller values for the short intervals, while the saving scores were considerably larger than the recall scores for the one- and two-day intervals.[8] The differences in the *size* of the two kinds of scores are thus not significant for our purposes.

It may be noted that the two methods do not give the same *order of retention* for the three materials for the abnormal subjects. From the standpoint of *relative* values, the relearning method has one serious defect. Due to its derivation, a saving score is a function not only of the amount retained but also of the rate of learning. Hence any difference in the saving scores for two lists will be determined in part by the difference in the learning scores for those two lists. The learning scores for the neutral material are much smaller than those for the other lists, and this fact has distorted the relative amounts retained as measured by the saving scores. The saving scores are an accurate measure of the relative retention of two lists

[8] Luh. The conditions of retention. *Psychol. Mono.*, 1922, 31.

only when their learning scores are approximately equal. The learning scores for the acceptable and unacceptable lists are approximately the same, and the two methods do give the same order of retention for these materials. The two methods thus give the same order of retention for those materials that are approximately equal in respect to ease of mastery. The acceptable materials are enhanced and the unacceptable materials are repressed relative to each other when retention is measured by either the saving method or the method of recall, and this statement is true for both classes of subject.

Conclusions

This experiment was designed to test Freud's doctrine of repression and enhancement (1) for materials that conform to Freud's usage of the terms pleasantness and unpleasantness, (2) for normal and abnormal subjects, and (3) for three intervals of time after learning.

The unpleasant or unacceptable materials exhibited a considerable amount of repression relative to the neutral materials. The repression was present two days after learning, and its amount did not materially change during the subsequent two weeks. The process of repression was thus limited to the first two days after learning.

The pleasant or acceptable materials also exhibited a small amount of initial repression, and a small amount of enhancement for the last two intervals. This change from repression to enhancement was due to an actual increase in the number of items recalled on the last two tests. This enhancement occurred during the two- to nine-day interval, as there was no evidence that the amount materially changed after this time.

The two types of subject exhibited no significant differences in respect to the phenomena of repression and enhancement. Both types of subject exhibited approximately the same amount of initial repression of the unpleasant items. Both exhibited practically the same amount of initial repression of the pleasant items, and both exhibited the same amount of the subsequent enhancement of these materials. The time during which these processes of repression and

enhancement occurred was the same for the two groups. Repression and enhancement are thus universal phenomena in that they are not limited to a certain type of subject. The essential similarity of the results for the two classes of subject may be cited in support of their reliability.

So far as these materials are concerned, the experimental results establish the truth of Freud's assumption of the existence of the phenomena of repression and enhancement. They also confirm his assumption of the universal character of these phenomena.

The data also reveal several unsuspected facts. (1) The process of repression affects both the pleasant and unpleasant materials but to an unequal degree. (2) In both cases, the process of repression is limited to the first two days after learning. This fact suggests that repression is an automatic process in that it takes place prior to the first recall. (3) Enhancement, on the other hand, is limited to the pleasant materials, and it does not occur until after the first recall. This fact suggests the possibility that the enhancement of the pleasant materials may be a function of the prior recall.

(Manuscript received December 13, 1937)

AN EXPERIMENT ON THE DISINHIBITION OF VOLUNTARY RESPONSES[1]

BY WALTER S. HUNTER

Brown University

Hudgins' experiment (**5**) on the conditioning of the pupillary light response to explicit and implicit verbal stimuli carries with it the suggestion that voluntary activity is essentially a conditioned response. In the final stages of that experiment, when the subject repeated the word which had been paired with an increase in light intensity, the pupil contracted. And when he repeated the word which had been paired with a decrease in light intensity, the pupil dilated. Thus the iris muscles were under the control of the verbal responses of the subject. Hunter and Hudgins (**6**) commented on the problem, in an interpretative way, as follows: "In conclusion we are of the opinion that the attempted classification of behavior into voluntary and non-voluntary forms of response is of less significance than a classification of the corresponding behavior in terms of genesis, type of control, and temporal characteristics. An adequate account of what the psychologist has called voluntary action cannot as yet be given because the necessary experimental data have not yet been secured; but the hypothesis is offered that so-called voluntary behavior is essentially a conditioned response having a characteristic latency and temporal course and under the control of self-excited receptor processes" (p. 118).

If we turn from Hudgins' experiment to the conditions of daily life, we find many of an individual's responses apparently under the control of his verbal mechanism. If he 'thinks' the words 'get up,' 'sit down,' 'move the finger,' or 'tap,' the specified reaction will often be made and will undoubtedly be entitled to rank as a voluntary response.

[1] This study was supported by a grant from the American Academy of Arts and Sciences.

However it also often happens that when an individual says or thinks, for example, 'move' with reference to his finger or 'wink' with reference to his eye-lid, the responses in question do *not* take place although it is known that the words or 'thoughts' and actions have been associated one with the other during the process of language acquisition. (My own belief is that the words and responses have been so related to each other by essentially conditioned reflex methods.)

Two reasons are suggested for the lack of invariability with which responses follow the presence of their appropriate words in the daily life of the individual: (1) In spite of the fact that the relevant responses often follow their appropriate words, these later are not the essential stimuli for the responses. The important cues are rather 'fiats of the will,' or the 'consent of the subject.' If the subject 'wills' that the finger make some response that through inheritance or learning it can make, the response will take place. The failure of the finger to obey indicates that the subject has not fully consented to or willed the movement. (2) A second possible explanation of the failure of, let us say, the finger to move or not to move when the appropriate words 'move finger' or 'don't move finger' occur may lie in the presence of inhibitions of the Pavlovian type. Some internal or external stimulus or some physiological state may be present which inhibits the response.

The first of the above explanations is unsatisfactory because there is no known method of testing its validity. In contrast to it, the second hypothesis is formulated at an experimental level, and it may be examined by well established conditioned response procedures. If failure to respond under the specified conditions is the result of inhibition, it should be possible to disinhibit the response by appropriate methods. The present experiment offers evidence in support of this second type of explanation.

Methods

Thirteen university students and one preparatory school student were used as subjects. No one of them knew the purpose of the experiment. None had served as subjects in previous conditioning experiments, although most of them were familiar with conditioning techniques in general.

During the experiment the subject was seated in a dark sound-reducing room. His right arm rested on a table and the middle finger of his right hand was placed under a tambor device of the type described by Watson (12). A metal rest on which the tip of the finger lay served as one shock electrode, and a brass plate under the palm of the hand was the other electrode. In some of the experiments a second tambor was placed under the finger, and its metal cover served as one electrode. Light stimuli could be given by either or both of two flash-light bulbs fastened to the wall about three feet in front of the subject and within 2 inches of each other. Just above these two lights was another concealed behind a cardboard screen but so adjusted that when turned on it would illuminate a card with a large printed word 'NO' on it. When not in use this card was turned face to the wall so that only a blank surface was exposed to the subject. An auditory stimulus was provided by a buzzer located about 18 ins. from the subject. A tactile stimulus could be given by a solenoid stimulator (Hull type) which was strapped to the subject's right fore-arm in certain of the experiments. The auditory and tactile stimuli were used solely as disinhibiting stimuli.

The experimenter and the remainder of the apparatus were in an adjoining room. The presentation of one light or two lights and the shock was controlled by contacts on a synchronous motor timer so arranged that approximately one second intervened between the conditioned and the reinforcing stimuli. The usual interval between such paired stimulations, or between one pair and a following unreinforced stimulus, was 59 secs., although frequent variations in this interval were made in order to prevent the subject depending upon a constant time interval. The buzzer, the tactile vibrator, and the light for the 'NO' card were controlled by manually operated switches. A telephone transmitter in the experimenter's room and a receiver (head-phone type) in the subject's room made it possible, when the procedure demanded it, for the experimenter to give the subject verbal instructions. These instructions were either 'Lift your finger' or 'Don't lift your finger.' When the telephone was not in use, its circuit was interrupted with a switch; and the head-phones were only worn during certain of the experiments. The presentation of each of the various stimuli and the finger reactions of the subject were recorded on a waxed paper kymograph driven by a constant speed motor.

Unless otherwise noted below, the subject was merely instructed to sit in a comfortable position with his finger resting on the tambor and his eyes open. The subjects all noted the possibility of shock, but they were in general told nothing about avoiding it. Preliminary tests were made to insure that the shock was adequate to produce finger movement but not too strong to be endured by the subject. It remained for the subject to learn that the shock could be avoided by lifting his finger promptly to the proper stimulus and keeping it lifted for a few seconds. Experimental sessions were one hour long with 3 min. rest periods at 15 min. intervals.

Results

Table I gives the experiments in which each subject served, the order of the experimental conditions to which he was subjected, and the positive or negative results of the disinhibition tests. One subject was tested under four conditions; four subjects, under three conditions; six subjects, under 2 conditions; and three subjects served in one experi-

TABLE I

General Summary of Results

Disinhibition of extinction:

Exper. 1 (col 2) condition to 1 light, extinguish, disinhibit, no instructions

Exper. 2. (col. 3) same as Exper. 1, except preliminary verbal instructions

Disinhibition of 2 lights:

Exper. 3 (col. 4) conditioned discrimination 1 light vs 2 lights, disinhibit, no instructions

Exper. 4 (col. 5) same as Exper. 3, except preliminary verbal instructions

Disinhibition of 'NO' card:

Exper. 5 (col. 6) conditioned discrimination 1 light vs 'NO' card followed by 1 light, disinhibit

Disinhibition of telephoned words:

Exper. 6 (col. 7) conditioned discrimination between telephoned commands plus 1 light, disinhibit

Entries show order in which subject took experiments and positive or negative results of disinhibition.

Subject	Experiment					
	1	2	3	4	5	6
A	1, +		2, +		3, o	
B	1, o		2, +		3, o	
C	1, +		2, +		3, +	
D			1, +		2, +	
E				1, +		2, +
F	1, +		2, +		4, +	3, +
G		1, +				2, +
H	1, +					2, o
I	1, +		3, +			2, o
J				1, +		2, +
K					1, +	
L						1, o
M			1, o			2, +
N				1, +		

ment only. Since disinhibiting stimuli lose their effectiveness on frequent repetition, it is impractical to test one subject many times for disinhibition with the same stimulus.

Disinhibition of Extinction.—In Experiment 1 six subjects were first trained to lift the finger in response to a single light which was always reinforced unless the subject responded promptly. No preliminary instructions about the light or about avoiding the shock were given. Conditioned responses were indicated by finger movements which followed the light but anticipated the shock. When five successive conditioned responses were made, extinction trials began and the rein-

forcing shock was never given. This continued until from three to five successive failures to respond to the light had been recorded. The sound of the buzzer was then presented *for the first time* 7–15 secs. before the light stimulus appeared again. The buzzer set off a definite startle response of the finger, but this had entirely ceased before the light was presented. In the one trial given each subject, the immediately following light elicited a finger response except with one subject. A similar disinhibition of the extinguished response was produced by the tactile vibrator in the one subject tested.

In Experiment 2 the one subject utilized had made no progress under the conditions of Experiment 1 where no instructions were given. He was therefore told, during a rest period, that if he would lift his finger when the light signal came he could avoid the shock. With this information he was soon able to respond to the level of the criterion mentioned above. The response was then extinguished. When the buzzer was presented for the first time, a marked startle response was given by the finger, and 15 secs. later a response was made to the light stimulus.

Disinhibition of Two Lights.—In Experiment 3 seven subjects were trained to discriminate one light (always reinforced unless the subject lifted his finger properly) from two lights. No preliminary instructions with reference to the visual stimuli or the finger movement were given. However, since all but two of the subjects had been utilized in previous experiments, it is fair to assume that they promptly instructed themselves about the method of avoiding the shock and sought only to discover which stimuli were being reinforced. After the discrimination was established, the startle finger response to the buzzer was reduced to zero by sounding the buzzer alone between the various light stimuli. The buzzer was then sounded simultaneously with the presentation of two lights. All subjects had at least one test, and two subjects received two tests each. In all cases but one, the complex stimulus produced a finger response. Figure 1, although it is concerned with another experiment, is equally illustrative of results in the present case. As Pavlov inter-

prets this experimental method, one light has an excitatory effect which is inhibited by the second light when two are presented. The disinhibiting stimulus serves to inhibit the inhibiting effect of the second light.

Experiment 4 utilized one subject (N) who had failed in 87 trials to discriminate one light from two lights under the conditions of Experiment 3 and two other subjects who were being tested for the first time. These subjects were told to lift their finger promptly for one light, thus avoiding shock, but not to lift the finger for two lights. In spite of these instructions, which were renewed during the rest periods, one subject (E) had great difficulty in not moving his finger to two lights. With the discrimination established and the startle finger response to the buzzer reduced to zero, a combination of two lights plus the buzzer produced finger responses in all subjects.

Disinhibition of the 'NO' Card.—Experiment 5 was planned to bring the language control of the finger response more clearly into prominence than may have been the case in Experiments 1–4. Six subjects in all were used. Table I shows the character of the previous training of these individuals, one of whom was serving his first session. Only the general instructions to sit quietly with eyes open and the finger in position under the tambor were given. The subjects were thus left to their own devices to learn that if the word 'NO' on the card was illuminated the immediately following single light would not be reinforced and the finger should not be moved. If the single light were presented without a preceding illumination of the 'NO' card, a shock would follow which could only be avoided by properly moving the finger. It was the experimenter's intention that the word 'NO' should in time come to inhibit the response which would otherwise be made to the light stimulus. Four subjects readily set up the discrimination; but the remaining two subjects had to be told the significance of the 'NO' card. After the discrimination was established, the startle finger response to the buzzer was reduced to zero. Then when the 'NO' card was presented followed by the complex stimulus of one light plus the buzzer, a finger response was elicited

from all four of the uninstructed subjects. The two subjects with whom the 'NO' card had been emphasized failed to show disinhibition. With these subjects, however, tests showed that two lights presented simultaneously aroused no response as opposed to a response to one light and that, as in Experiment 3, the buzzer (reduced to zero) would produce disinhibition when presented simultaneously with the two lights.

Disinhibition of Telephoned Words.—Experiment 6 increased the emphasis upon the verbal control of the subject's finger. Only one light stimulus was used. When its appearance to the subject was preceeded by the telephoned words 'Lift your finger,' the light was reinforced; and when it was preceded by the words 'Don't lift your finger,' no shock was given. Eight subjects were tested under these conditions, one person for the first time. After the subjects had received from 50–75 trials and were reacting consistently, the startle finger response to the buzzer and to the tactile vibrator were reduced to zero. Disinhibition tests were then made by saying over the telephone 'Don't lift your finger' and then presenting either light plus buzzer or light plus tactile vibration. Five of the subjects showed disinhibited finger responses under these conditions while negative results were secured from three subjects. Figure 1 shows a sample record

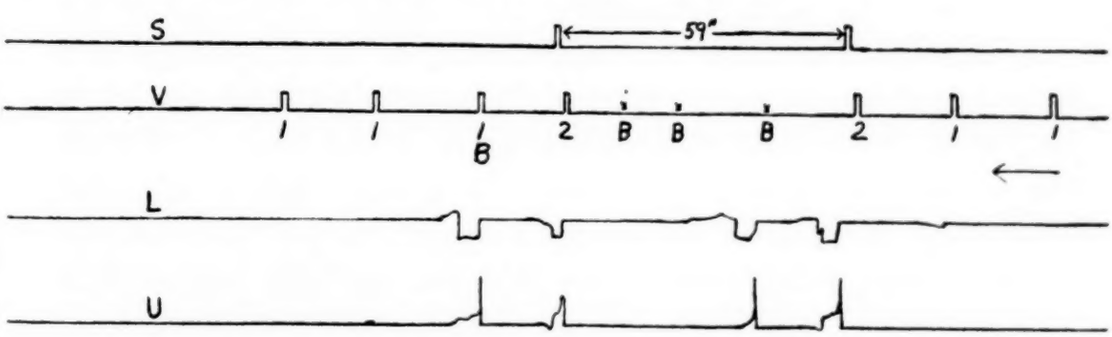


FIG. 1. Disinhibition of telephoned commands. Subject *J. S*, shock; *V*, presentation of one light; *B*, buzzer; 1, verbal stimulus 'Don't lift your finger' given just before the light; 2, verbal stimulus 'Lift your finger' given just before the light; *L*, lower tambor record of finger movement; *U*, upper tambor record of finger movement. (Read from right to left.) With the startle response to the buzzer extinguished, the next light stimulus was preceded by the words 'Lift your finger,' and the response was made. (Shock was present, but avoided.) The next stimulus was 'Don't lift your finger' immediately followed by a simultaneous presentation of light and buzzer. The result was a disinhibition of the finger response.

of disinhibition with one of the subjects. There is no difference between the type of phenomenon found here, where a verbal inhibition of response is inhibited, and that found in the earlier experiments where the inhibiting effect of a second light is itself inhibited by a buzzer sound.

Interpretative Comments

The results of the present experiments are in harmony with an increasingly significant body of data gathered by other investigators and tending to show the importance of verbal processes in human conditioning. Some of this earlier work may well be mentioned here. In 1919, three years after Watson's first paper (**12**), I. A. Hamel (**4**) presented results which indicated the similarity between conditioned and voluntary responses in latent time and which were interpreted to emphasize the role of 'consciousness' in human conditioning. In 1929 Schilder (**10**) also stressed the involvement of 'psychic' processes; and in 1933 Hudgins (**5**) published his important demonstration of the verbal conditioning of the pupillary light reflex. In 1934 and 1935 the role of verbal activities in conditioning was further indicated by Cason (**1**, **2**). Since then Cook and Harris (**3**, 1937) have verbally conditioned the galvanic skin reflex, and Menzies (**7**, 1937) has secured verbal conditioning of vasomotor responses. In 1932 Schlosberg (**11**) called attention to the irregularities found in the conditioning of adult human subjects and wrote in part as follows: "Thus the conditioned knee-jerk seems to be a bit of behavior more or less broken off from our usually integrated patterns. In a way, this is what we would expect when we encourage a subject to pay no attention to the conditioning process, or at least ask him not to kick 'voluntarily.' . . . This is markedly different from the methods used with young children and lower animals, where every effort is made to avoid distracting stimuli, and to focus the organism on the stimuli and responses involved in the conditioning. But if this method were adopted with adult adult human subjects, we should be forced to classify the experiment under some heading other than that of 'condi-

tioned response'" (p. 337). Razran (**8**) in 1935 and again (**9**) in 1936 presented an extensive study of conditioning the adult human salivary response in which he stressed attitudinal controls as responsible for the differences between human and animal conditioning and extinction. He wrote in part as follows: "When a consciously reportable and consciously controllable response is being conditioned, the subject assumes, after a few presentations, of the conditioned and conditioning stimuli, a certain attitude or attitudes towards the conditioning situation. These attitudes are best conceivable in terms of some postural set or some verbal organization . . . (pp. 41–42).

In the present study, completed in 1934–5, I have attempted a further development of this point of view through the utilization of disinhibition. If a verbal stimulus, like the 'NO' card or the telephoned command 'Don't lift your finger,' is exerting an inhibitory effect upon the normal excitatory stimulation from the light, it should be possible to inhibit the inhibition and so bring forth the response. This we have done. Furthermore in Experiments 2 and 4 the subjects who were given definite verbal instructions as aids in the conditioning manifested the same disinhibition as those subjects who were not so aided in Experiments 1 and 3. It seems reasonable to assume that the subjects were always working under verbal instructions which they either gave themselves or which came from the experimenter and that therefore all of the conditioned responses were to that extent voluntary. As Cason (**2**) says, "In the conditioned hand withdrawal experiment with a person as the subject, the person's verbal activities will be present, and they will be active in a different way at different times, regardless of whether or not he is told that he is supposed to learn something and regardless of whether or not he is told what it is that he is supposed to learn" (p. 317). It is impossible to put an intact human subject into an experimental situation and not have that subject try to solve the problem with which he is confronted, although the problem as formulated by the subject may not be the same as the problem formulated by the

experimenter. In such a subject verbal activities will always be effective in controlling some part of the behavior, and, as a result of training, the verbal processes may finally control the particular responses which the experiment is designed to study. Such responses should be ranked as 'voluntary' ones since any behavior which is brought under the control of processes that can be excited by the subject will have the characteristics of apparent spontaneity, control by the subject, and long latency which typify voluntary activity.

The disinhibition studies of the present paper contribute data in support of such a view of the nature of voluntary responses by showing that these activities are susceptible to the types of disinhibition which Pavlov described for conditioned reflexes. The results further support the view that when customary responses fail to follow the arousal of their appropriate verbal stimuli, other stimuli being constant, the explanation lies in an inhibitory process which under favorable circumstances can itself be blocked and the original response disinhibited.

(Manuscript received December 28, 1937)

1. Cason, H., The role of verbal activities in the conditioning of human subjects, *Psychol. Rev.*, 1934, **41**, 563–571.
2. Cason, H., An attempt to condition hand withdrawal responses in human subjects, *J. Exper. Psychol.*, 1935, **18**, 307–317.
3. Cook, S. W., and Harris, R. E., The verbal conditioning of the galvanic skin reflex, *J. Exper. Psychol.*, 1937, **21**, 202–210.
4. Hamel, I. A., A study and analysis of the conditioned reflex, *Psychol. Monog.*, 1919, **27**, No. 118.
5. Hudgins, C. V., Conditioning and the voluntary control of the pupillary light reflex, *J. Gen. Psychol.*, 1933, **8**, 3–51.
6. Hunter, W. S., and Hudgins, C. V., Voluntary activity from the standpoint of behaviorism, *J. Gen. Psychol.*, 1934, **10**, 198–204; also in *Acta Psychol.*, 1935, **1**, 111–118.
7. Menzies, R., Conditioned vasomotor responses in human subjects, *J. Psychol.*, 1937, **4**, 75–120.
8. Razran, G. H. S., Conditioned responses, *Arch. of Psychol.*, 1935, No. **191**.
9. ——, Attitudinal control of human conditioning, *J. Psychol.*, 1936, **2**, 327–337.
10. Schilder, P., Conditioned reflexes, *Arch. Neurol. & Psychiat.*, 1929, **22**, 425–443.
11. Schlosberg, H., An investigation of certain factors related to ease of conditioning, *J. Gen. Psychol.*, 1932, **7**, 328–342.
12. Watson, J. B., The conditioned reflex in psychology, *Psychol. Rev.*, 1916, **23**, 89–116.

FATIGUE OF THE VIBRATORY SENSE[1]

BY C. H. WEDELL AND S. B. CUMMINGS, JR.

Princeton University

Although the perception of mechanical vibration through the skin is a common experience, it has only recently received attention from the experimental psychologist. The lack of scientific interest was probably due to the fact that although the feeling of vibration is of everyday occurrence, it does not seem to play a very important part in the adjustment of the human being to his environment. Not until it was realized that a knowledge of the sensory mechanism involved in vibratory stimulation of the skin would contribute greatly to the understanding of the other cutaneous sensory functions did investigations of this problem become numerous.

These investigations have resulted in a considerable number of facts concerning vibratory sensitivity. Determinations have been made of the absolute and differential limens for both frequency and intensity (**3**, **6**) and of other psychophysical functions (**2**, **7**, **8**). However many other characteristics of the vibratory sense must be known before its nature can be adequately described, and one of these characteristics is fatiguability. The experiment to be reported in this paper is a study of the effect on the threshold of vibration of prolonged stimulation at various frequencies and intensities.

Only two experimenters have attacked this problem directly. Kampik (**4**) had his subjects clutch a wooden stick which was set into vibration by means of a motor-driven toothed wheel. He found that at the end of an hour the perception of vibration had disappeared and the subject complained that his arm felt numb for some time afterward. In this experiment no attempt was made to relate the amount of fatigue to the characteristics of the stimulus.

[1] From the Psychological Laboratory of Princeton University.

The other experiment is the recently published study by Bellows (1). He used as a stimulus an interrupted air-blast whose frequency, pressure and temperature could be accurately controlled. It is somewhat difficult to determine the amount of adaptation [2] from his results since his method was that of the critical frequency of flicker. The subject was stimulated by an interrupted air-blast at a certain pressure and had to adjust the frequency himself to maintain the stimulus just below the fusion point. By rearranging Bellows' data, it is possible to show that in order to maintain the three frequencies 68, 85 and 105 cycles at the critical fusion point, it was necessary to increase the pressure gradually until it had doubled at the end of 14 minutes of continuous stimulation. This amounts to a loss of sensitivity of 6 db.

It should also be noted that Katz (5) states that the vibratory sense does not fatigue and resembles the auditory sense in this respect. However he does not offer any experimental evidence in support of his contention.

Apparatus and Procedure

The frequencies used in the investigation were 64, 256, 700 and 1024 cycles per second. They were electrically generated by a Western Electric 2A audiometer which had been slightly modified to make possible the production of 700 cycles. The currents were led from this audiometer through a voltage attenuator graduated in steps of 1 decibel, to a vacuum-tube, power amplifier which drove a magnetic loud-speaker motor. A short, blunt bakelite rod, 20 mm long and 3 mm in diameter, was screwed to the armature of the motor. The end of the rod was shaped into a rounded point which made contact with the skin of the subject over an area 1 mm in diameter.

The subject sat in a specially constructed chair with the lower right arm and hand resting palmar side uppermost in a plaster-cast form. The arm was not held by straps and the subjects were able to maintain an easy, relaxed position without discomfort for the duration of each experimental session. The vibrator was lowered onto the palm of the hand so that the fiber rod made a slight depression in the skin at a point on line with the axis of the third finger, 1½ inches from the wrist. Care was taken to see that the subject could feel a firm yet not painful pressure.

At the beginning of the experiment, each subject's threshold was determined by the method of limits. Twenty judgments were made at each frequency. All intensity values in the experiment are expressed in decibels above or below this threshold value.

---

[2] Bellows uses the term 'adaptation' and not 'fatigue.' There is some confusion in the psychological literature in the use of these terms. The term 'fatigue' is used in this paper because it seems more logical to restrict the application of the term 'adaptation' to those changes involving a better adjustment of an organ or the whole organism to its environment.

The possibility of variation in the threshold during the course of the experiment was checked by taking four readings at the beginning of each session. If these readings showed that the threshold had changed by more than 3 db, another series of 20 judgments was made and the average of these was used as the new threshold value. Two thresholds showed a change of 5 db and one a change of 10 db during the experiment. All of these changes were in the direction of greater sensitivity.

The procedure used throughout was to stimulate for 3 minutes with a frequency of 64, 256, 700 or 1024 cycles per second at an intensity of 10, 25 or 40 db above threshold. Immediately following the cessation of stimulation, 5 pairs of threshold measurements were made. A pair of judgments consists of one judgment made with the tone increasing in intensity from well below threshold and one as it decreased in intensity from a point above threshold. These 10 readings took about 2 minutes. About 1 min. later or three minutes after the end of the fatiguing stimulation, two more pairs of readings were made and again at 5, 7, and 9 minutes after the end of the fatiguing tone. The tone used as the fatiguing frequency is called the 'stimulation' or 'fatiguing' tone and the frequency upon which tests were made following the fatiguing period is called the 'test' tone or frequency. Since 1024 cycles was audible before it was felt as vibration, the noise from an electric fan was used as a masking tone whenever this frequency was used.

The 7 subjects were all male graduate students or instructors in the department of psychology. One of the subjects was used only in the check experiment to test the effect of overtones, as described below, and the results from the other six form the main body of data. The experimental sessions were spaced over a period of weeks with two sessions ordinarily being given in the same hour, separated by a rest period.

Results

Amount of Fatigue and its Relationship to Frequency.— Three frequencies, 64, 256 and 1024 cycles per second were used as fatiguing tones at an intensity level of 25 db above threshold. Each of 6 subjects served for three sessions in which the test tone was the same as the fatiguing tone. Figure 1 shows the results obtained. In this figure, the abscissa represents time in minutes from the cessation of the fatiguing tone, while the ordinate shows the loss of sensitivity in decibels below the normal threshold. Each point on the curves is the average of 12 pairs of threshold judgments, one pair on each of the six subjects.

These results show that fatigue of the vibratory sense does occur under the conditions of the experiment. The amount of fatigue depends somewhat upon the frequency of stimulation. Immediately after stimulation, the loss of sensitivity is 10 db at 64 cycles, 14.5 db at 256 cycles and 15 db at 1024 cycles. There follows a gradual recovery until at the end of

9½ minutes, normal sensitivity to 64 and 256 cycles has returned and 1024 cycles is within 3 db of normal.

The difference between the curves for 64 and 1024 cycles probably represents a real difference between the effects of stimulation at these two frequencies since it appears in the records of 5 of the 6 subjects.[3] In the sixth subject, all three curves are nearly identical. On the other hand, the differences between the curve for 256 cycles and the other two

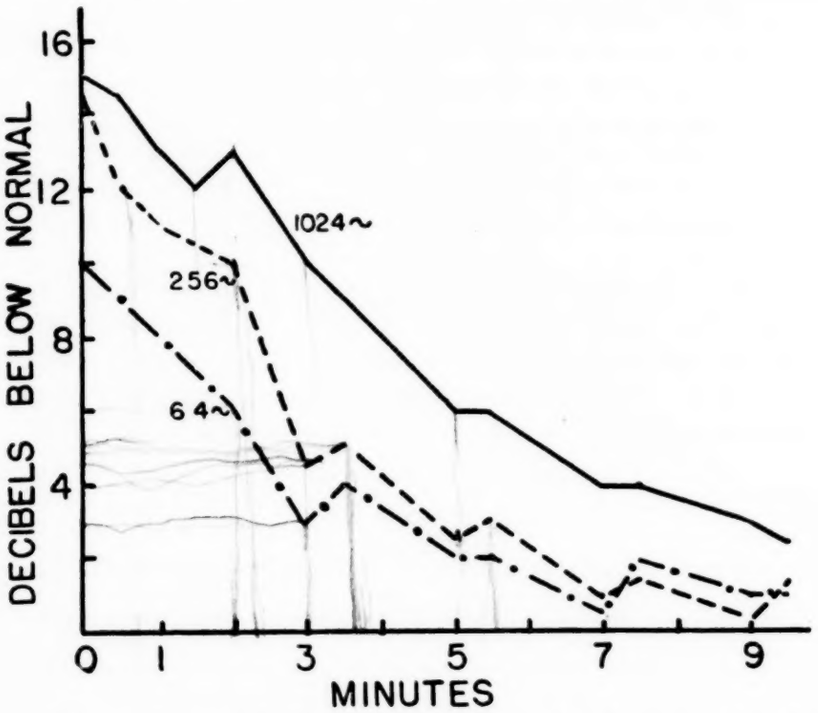


FIG. 1. Return to normal of sensitivity to 64, 256 and 1024 cycles after 3 min. continuous stimulation with the same frequency at 25 db above threshold. The abscissa shows the time in minutes after the cessation of stimulation. The ordinate shows the loss of sensitivity in terms of decibels above the normal threshold. Each point on the curves is the average of 12 judgments.

curves are not reliable as judged by the fact that in some subjects the curve for 256 cycles is nearly identical with that of 64 cycles while in others it shows almost the same fatigue effect as 1024 cycles. These facts indicate that frequency has an effect although it is small. More data are needed to determine whether the relationship between amount of fatigue and frequency is direct or whether a more complex situation exists.

[3] This statement is based upon an examination of the records of the individual subjects. Only the averages of all subjects are presented in this paper.

Loss of Sensitivity in Relation to the Intensity of Stimulation. —The same 6 subjects served for two more experimental sessions in which 256 cycles was used as the fatiguing tone at intensities of 10 db and 40 db above threshold. Figure 2 shows the two curves obtained in comparison with the curve for 256 cycles at 25 db taken from Fig. 1. The ordinate and abscissa are the same as in Fig. 1. It is clear from Fig. 2 that the greater the intensity the greater the fatigue effect,

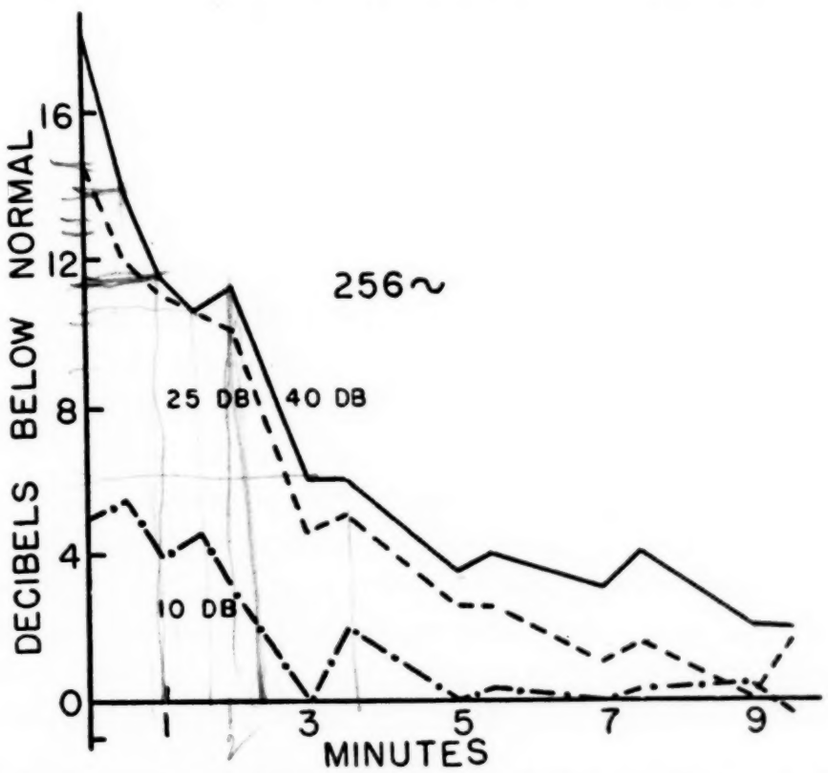


FIG. 2. Recovery of sensitivity to 256 cycles after 3 min. continuous stimulation with this frequency at three different intensity levels, 10, 25 and 40 db above threshold. Abscissa and ordinate as in Fig. 1.

at least for 256 cycles. When the fatiguing tone was only 10 db above threshold, as represented by the dot-dash line of Fig. 2, the loss of sensitivity immediately after stimulation was only 5 db. Within 5 minutes, the effect of the continuous stimulation had disappeared. However with a fatiguing intensity of 40 db, the immediate loss of sensitivity was 18 db as seen in the solid line curve and sensitivity was still 2 db below normal after 9.5 minutes.

The difference between the curve for 10 db on the one hand and the curves for 25 db and 40 db on the other is a significant

difference since it was shown by all six subjects. Furthermore, in all six subjects, the curve for 25 db is nearer the curve for 40 db than the curve for 10 db. In some subjects, the curves for 25 and 40 db are nearly identical, while in others the curve for 25 db is slightly below that for 40 db. This indicates that after stimulation for three minutes with 256 cycles at 25 db above threshold, the loss of sensitivity is nearly maximum and further increase in intensity does not result in any measurably increased fatigue.

Specificity of the Fatigue Effect.—It is of interest to know whether stimulation at one frequency affects all other frequencies or whether the loss of sensitivity is restricted to the fatiguing tone. To determine this question, the same subjects served for six sessions in which the fatiguing tone and test tone were of different frequency. The tones used, namely 64, 256 and 1024 cycles, were always 25 db above threshold. Combining the results obtained with those in the first part of the experiment where the stimulation tone and test tone were the same, we have the nine curves shown in Fig. 3. The ordinate and abscissa are the same as before. Each group of three curves shows the fatigue effect on all three frequencies resulting from stimulation at one frequency. It is evident in this figure that a fatiguing tone of 64 cycles affects all three frequencies equally. However, 256 cycles affects itself and 1024 cycles equally but does not have as great an effect on 64 cycles. Finally 1024 cycles, the highest frequency, affects both 256 and 64 cycles less than 1024 cycles itself. In general terms, this means that a frequency A affects a frequency B as much as A is affected by itself, provided B is higher than A. If B is lower than A, then B will not be affected by A as greatly as A fatigues itself.

On the other hand, while a given tone does not affect a lower tone as much as it does a higher one, the given tone may still affect the lower tone as much as the latter would affect itself. This fact is illustrated in the middle group of curves in Fig. 3. The tone of 64 cycles does not show the effect of stimulation at 256 as much as do 256 and 1024 as already noted but when comparison is made with the first

group of curves, 64 cycles is seen to be affected by 256 as much as by 64 cycles.

Check Experiments.—No filter was used in the tone-producing apparatus and overtones were undoubtedly present. The effect of one frequency on higher frequencies may be due to the presence of these overtones. In order to test this possibility, three of the subjects together with one new subject served in 4 experimental sessions in which 256 and 700 cycles were used in all possible combinations of stimulation and test

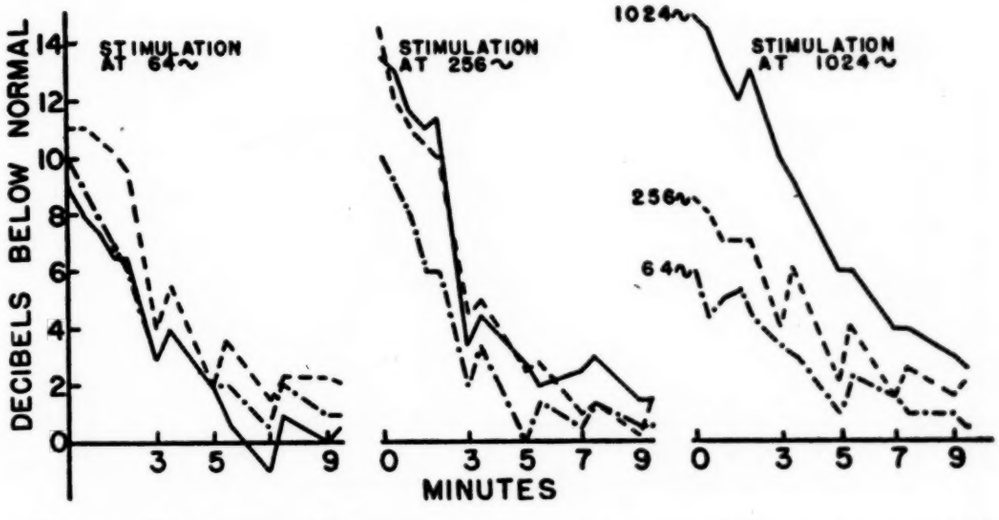


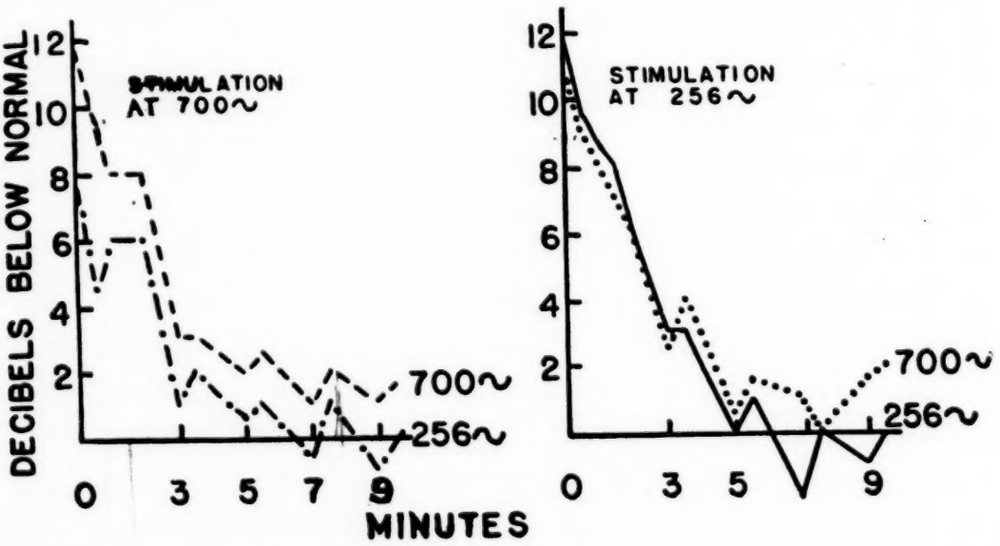
FIG. 3. Each group of three curves shows the course of recovery of sensitivity to 64 cycles (dot-dash line), 256 cycles (dash line) and 1024 cycles (solid line) after stimulation for three minutes at 25 db above threshold by the frequency indicated over that group of curves. Abscissa and ordinate as in Fig. 1.

frequency. 700 cycles was chosen because it is not an overtone of 256 cycles. The data are shown in Figs. 4 and 5.

If the conclusions deduced from Fig. 3 are to hold true for 256 and 700 cycles, the curve showing the result of stimulation at 700 cycles and test at 256 should be lower, *i.e.* show a smaller loss of sensitivity than the curve resulting from stimulation at 700 cycles and test at the same frequency. It can be seen in Fig. 4 that this is true. In this figure, the dash line gives the result of stimulation at 700 and test at 700 cycles. The dot-dash line represents the data obtained from stimulation at 700 cycles and test at 256. The latter is everywhere below the former, showing that after stimulation at 700 cycles, 256 shows a smaller loss of sensitivity than 700

cycles. Although the difference between the two curves is not great, all four subjects are consistent in showing the same effect.

Furthermore, if the results obtained with 256 and 700 cycles are to be consistent with those obtained with the other frequencies not only should one curve in Fig. 4 be below the other, but 256 cycles should have as great an effect on 700 cycles as on itself. Figure 5 shows that this is true. The solid line represents the effect of stimulation at 256 cycles on the threshold for the same tone. The dotted line represents



FIGS. 4 and 5. Figure 4 shows the change in threshold for 700 cycles and 256 cycles following fatiguing stimulation for three minutes with 700 cycles at 25 db above threshold. Figure 5 is the same except that the fatiguing tone was 256 cycles. Ordinates and abscissae are the same as in the previous figures. Each point on the curves is the average of eight judgments.

the effect of the same fatiguing tone on the threshold for 700 cycles. The curves are remarkably similar and this is the case for all four subjects. The averages plotted in the curves of Fig. 5 do not conceal any significant individual differences. This check experiment shows clearly that the effect of a low tone on a higher one is not an artifact due to the presence of overtones. It was thought possible that simple adaptation of the touch receptors under the stimulation point might be the cause of part of the fatigue obtained although it would not help to explain the differential effect just described. To

eliminate this possibility, the fiber rod was allowed to rest quietly on the skin in the normal stimulating position for three minutes. At the end of this time, the threshold measurements showed no change whatever in sensitivity.

Introspective Data.—Although no attempt was made to obtain systematic introspections on the quality and intensity of the conscious experience aroused by the continuous stimulation, one or two spontaneous comments by the subjects are worthy of mention. In the first place, the feeling of vibration had never disappeared at the end of the three minute period. Whenever the subject wished to attend to the sensation, it was always there but several of them reported that it had decreased in intensity by the end of that time.

In the second place, it is interesting to note that the vibration was not always localized under the point of stimulation. It was quite common for the observer to report that the feeling of vibration seemed to be spread over the hand or part of the arm. However, as the intensity of stimulation was increased from below threshold, the sensation always appeared at a point which was characteristic of the subject and was constant from session to session.

Summary

1. Sensitivity to vibratory stimulation applied to the palm of the hand is reduced 5 to 15 decibels after three minutes continuous stimulation.

2. The loss of sensitivity is greater the higher the frequency and the greater the intensity of the fatiguing tone.

3. After stimulation for three minutes at a certain frequency, the loss of sensitivity is the same whether measured at a frequency equal to or higher than the fatiguing frequency, but it is less if measured at a lower frequency.

(Manuscript received December 11, 1937)

References

1. Bellows, R. M., An experimental isolation of some factors determining response to rhythmic cutaneous stimulation. I. Frequency, pressure and time, *J. Exper. Psychol.*, 1936, **19**, 716–731.

2. Fessard, A., Le problème des excitations tactiles brèves, *Année Psychol.*, 1928, **29**, 142–165.
3. Gilmer, B. von H., The measurement of the sensitivity of the skin to mechanical vibration, *J. Gen. Psychol.*, 1935, **13**, 42–61.
4. Kampik, A., Untersuchungen über die Leistungsfähigkeit der Vibrationsempfindungen, *Arch. f. d. ges. Psychol.*, 1930, **76**, 3–70.
5. Katz, D., The vibratory sense and other lectures, *U. of Maine Studies*, Orono, Maine, 1930, 90–105.
6. Knudsen, V. O., 'Hearing' with the sense of touch, *J. Gen. Psychol.*, 1928, **1**, 320–352.
7. Newman, H. W. and Corbin, K. B., Quantitative determination of vibratory sensibility, *Proc. soc. exper. biol. and med.*, 1936, **35**, 273–276.
8. Pearson, G. H. J., The effect of age on vibratory sensibility, *Arch. neurol. and psychiat.*, 1928, **20**, 482–496.

WHOLE AND FOUR PART LEARNING THIRTY TWO UNIT SPIDER MAZES

BY THOMAS W. COOK

Acadia University

I. Problem and Procedure

An earlier paper (1) reported data dealing with the relative economy of whole and part methods of learning 'spider' mazes assembled from standard units with entrance, exit, and 6 culs-de-sac in each unit. In that investigation each of 3 subjects learned 38 patterns of 4 degrees of complexity: 2, 4, 8, and 16 units. For part procedure a 2-part division was used. The present paper describes a study of the same subjects learning similar maze patterns, but differing from the earlier experiment in that the 'whole' patterns were made up of 32 units, and in part procedure each of *four* 8 unit parts were first learned separately and afterwards traced 'as a whole' (pure part method). Each subject learned 16 mazes, 8 by the whole method and 8 by the part method. The order of learning by part or whole procedure was wppwwppwwp-pwwppw for subject W, and pwwppwwppwwppwwp for subjects M and C. Twenty-five seconds elapsed between trials, and the mazes were learned for the most part one every second day. In a few instances there were unavoidable variations from this arrangement, but in no case was the interval less than a day. As before, the criterion of learning was 3 errorless trials, except that with subject C (because of the extreme difficulty that she found in deleting the last few errors with 32 unit patterns) after maze 4 only one errorless trial was required in whole learning or combining the parts. The only change in the apparatus, other than increasing the number of blocks, was the use of two screws through each 4 inch block for holding the units in position. Otherwise the procedure was identical with that of the previous experiment.[1]

[1] In both experiments C was paid a small sum of money for each maze learned.

II. Results

A. First Trial

(1) *Part-whole Comparison.*—On trial 1 part-whole differences were negligible for subjects M and C. M's results show an average time of 137 seconds and 80.8 errors with the part method and 140 seconds and 78.6 errors with the whole method; C required 215 seconds and made 81.5 errors by the part method and 224 seconds and 82.0 errors by the whole method. Subject W, however, took more time for part than for whole tracing (198 to 175 seconds) and made fewer errors (75.2 to 85.5). The time record differences between the 3 subjects are due to individual variations in method of tracing. C moved slowly and cautiously. Her fashion of trying the paths (exit or cul-de-sac) in irregular order did not favor speed. W and M entered the blind alleys in regular order from entrance to exit in each unit, but M had by now reduced this to a routine to be got through as fast as possible, and indeed showed a small but constant improvement in speed for each successively learned group of 4 mazes. W, on the other hand, moved more slowly and counted aloud numbers corresponding to the serial order (1, 2, 3, etc.) of staples from entrance to exit on a given block. Then if she had taken the shortest path she had only to note the number of the exit, or, if she had chanced on the longer route, to subtract that number (5, 6, or 7) from the total number of staples on a unit (with a mental reservation that the alternate route around the central block should be taken on the next trial) in order to find and verbalize a number corresponding to the shortest path from entrance to exit on any particular unit. It should be noted that W counted the exit as well as the preceding blinds. Thus the 'number' for a given unit would be '1' if the true path were immediately adjacent to the entrance, and '4' if the true path were as far as possible (180°) from the entrance.

The only additional task incident to complete mastery of a maze (except learning whether to turn left or right on entering each unit) was memorization of these numbers in the order in which the corresponding units occured in the

pattern. Such memorization would ordinarily receive no special emphasis during the first trial. But in the present experiment W often paused at the entrance to a unit to rehearse the order of numbers corresponding to the shortest path over each unit from the beginning of the maze. This undoubtedly accounts for her slower time in part procedure, since rehearsal proved less feasible for 32 unit patterns than for 8 unit patterns, being restricted in the former case to the first few units in the pattern.

W's larger number of errors in whole procedure is more difficult to explain. It could only be due to choice of the shorter way around the central block on each unit more frequently in part than in whole procedure, since both M and W entered all blinds on whichever of the two alternate routes they chose to take. As noted in the previous paper (1), some factor in the maze situation makes the number of errors less than it would be if this selection were a matter of chance. Chance expectation is 96 (32×3) errors per maze, whereas the average for all 3 subjects is slightly over 80. But I can see no reason why such a factor should operate more effectively in part than in whole procedure. Of course, the fact that W learned different mazes in part and in whole procedure may be to some extent responsible for the differences in first trial errors with the two methods. Subject M's results show a slight trend in the opposite direction, and since M and W used whole and part procedure in counterbalanced order, their combined results would be free from the influence of characteristics of specific maze patterns. Such a combination shows an average of 82 errors for whole procedure and 78 errors for part procedure, a difference of only 5 percent.

(2) *Practice.*—First trial errors are little influenced by practice within the limits of the experiment. The averages for all 3 subjects with mazes 1–4, 5–8, 9–12, 13–16 are 82.0, 80.0, 82.7, 77.6. Time shows some slight practice effect, but the decrease is small and variable except with subject M. The latter's average time for each successively learned group of 4 mazes is 151, 144, 140, 114 seconds.

(3) *Regional Errors.*—The average errors are about the same for each of the 4 parts of the mazes. These averages for whole and part procedure and all 3 subjects are: parts I, II, III, IV, 20.8, 19.8, 19.3, 20.7.

(4) *Size and Difficulty.*—For both time and errors difficulty is directly proportional to size of pattern with subjects M and C. With W difficulty increases slightly less rapidly (3.5) for time and slightly more rapidly (4.5) for errors than the difference in size (4 times) between the 8 unit and the 32 unit patterns.

B. Later Trials

(1) *The Relative Economy of Whole and Part Procedure.*—For all trials except the first the part method is superior on the average to the whole method (Table I) 165 percent in

TABLE I

WHOLE VERSUS PART LEARNING. TIME IN SECONDS

Subject	Errors			Time			Trials		
	W	M	C	W	M	C	W	M	C
Whole	210.8	222.6	243.6	1313	1505	1529	9.0	11.9	9.9
Part	66.0	77.4	129.5	1048	1210	1839	9.6	10.4	12.1
Percent super.									
of part (totals)	219	187	89	25	25	−20	−7	15	−22
of First half	223	199	82	27	20	−26	−6	9	−40
of Second half	214	172	96	23	30	−14	−8	22	− 4

errors and 10 percent in time, but is 5 percent less economical in number of trials. These findings agree in the main with those of the earlier experiment using spider mazes and the same subjects. The individual averages for M and W (the two adults) are remarkably alike both for the part method and for the whole method, but the likeness, as far as the part method is concerned, is due to the 'accident' that W's superiority in initial part learning is balanced by M's greater efficiency in combining the parts. Subject C showed less advantage for the part method in errors (89 percent) and a slight advantage for the whole method in time (20 percent). It is worth noting that C was almost as efficient in whole

learning as M or W. The latter, however, were much more efficient than C in part learning.

There was some improvement through practice from earlier to later stages of the experiment, but Table I indicates that this had little or no influence on the relative economy of whole and part procedure. The part method is 223, 199, and 82 percent more economical of errors with W, M, and C for mazes 1–8, and 214, 172, and 96 percent more efficient with mazes 9–16. Time and trials show like results. The only exception is subject C's record of 40 percent advantage for the whole method by the 'trial' criterion with mazes 1–8, as compared with only 4 percent advantage with mazes 9–16. It is possible that this represents a gain for part procedure with practice.

(2) *Amount of Material*

(*a*) *Relative Difficulty of Parts I, II, III, IV.*—The average errors in initial part learning were 14.3, 15.9, 10.5, and 8.8 for parts I, II, III, and IV respectively. At first sight these values seem to indicate constant 'regional' differences, but part of the variation is probably to be ascribed to differences in difficulty between various sections of the patterns or to other 'chance' factors. For example, if the first 2 mazes learned by M and the first 3 learned by C are not included, the averages become 11.9, 9.5, 7.3, 9.4. It is possible, however, that with 4 part learning factors inherent in the method (such as warming-up or transfer) obscure to some extent the relation between amount of material and difficulty of learning.

(*b*) *Relation of Size of Pattern to the Whole-part Problem.* (*x*) *Initial Part Learning and Whole Learning.*—The relation between length and difficulty is shown in Table II. The results for each subject are given separately, as W's data differ strikingly (in degree) from those of other two subjects. When trial ratios are converted into 'units presented' by multiplying the number of trials for whole learning by 4 (**2**, p. 16), M's and C's results indicate that with this measure as well as with errors and time difficulty increases slightly less than 'as the square of' size of maze pattern. The square of

TABLE II

LENGTH AND DIFFICULTY. TIME IN SECONDS

	Subject			Subject M			Subject C		
	8 unit	32 unit	Ratio	8 unit	32 unit	Ratio	8 unit	32 unit	Ratio
Errors	4.3	210.8	49	15.4	222.6	15	17.4	243.6	14
Time	46	1313	29	109	1505	14	125	1529	12
Trials	1.5	9	6	3.4	11.9	3.5	3.3	9.9	3

the increase in size from 8 to 32 unit patterns is 4 × 4, or 16. The actual increases in difficulty for errors, time, and units presented are 15, 14, and 14 (3.5 × 4) for M and 14, 12, and 12 for C. These ratios for M and C are probably slightly lower than they should be, since there was a considerable improvement with practice from the first to the second maze with all 3 subjects, and both M and C learned maze 1 by the part method.

Since subject W learned maze 1 by the whole method the initial practice effect would make her ratios slightly higher than would otherwise be the case. Even if we omit maze 1, however, the difference in difficulty between 8 and 32 unit patterns is much greater for W than for M or C. With W the 32 unit patterns yield on the average 45 times (49 in maze 1 is included) as many errors, and require 26 times as many seconds and 22 times as many units presented (5.5 × 4) as the 8 unit patterns. The major discrepancy between M and C on the one hand and W on the other is thus in the error measure for 8 unit learning. W is only slightly more efficient than M or C with 32 unit patterns, and the most plausible interpretation of W's superior time and trial records in 8 unit learning is that they are a secondary consequence of rapid error elimination. It seems probable, therefore, that some feature of W's technique was particularly effective for deleting errors with the 8 unit mazes, but of little aid with the 32 unit patterns. Observation of W's behavior supplemented by verbal report confirms this interpretation, and identifies the factor as the rehearsal (discussed above) during trial 1, at the entrance to several units, of 'the order of numbers

corresponding to the shortest path over each block from the beginning of the maze.' Such review of recently traversed sections of the maze was considered responsible for W's slightly lower relative speed in first trial tracing of 8 as compared with 32 unit patterns, since in the latter case W was able to review only the first few units of the total pattern. Apparently, however, the additional time spent in first trial tracing effected a great error economy in subsequent trials with a given 8 unit pattern.

Neither M nor C used the 'rehearsal' method, although according to reports given at the end of the experiment M counted (silently) often and C occasionally.

(*y*) *Combining the Parts.*—Subject W had relatively great difficulty in integrating the parts after separate learning. She made an average of 48.9 errors, 633 seconds, and 5.1 trials in combining, or (excluding the first trial and the last 3 trials in order to make error, time, and trial measures comparable) 74, 77, and 77 percent of the totals for the part method. W therefore expended about 3 times as much effort in uniting the parts as in learning them in the first place. Moreover, her range of records for the combining act show a normal (chance) distribution and no practice effect. For example, the number of errors made in combining the parts of mazes 2, 3, 6, 7, 10, 11, 14, 15, were 45, 72, 10, 89, 21, 35, 61, 58 respectively.

Subject M made fewer errors in integrating, and showed considerable improvement from practice. His errors for mazes 1, 4, 5, 8, 9, 12, 13, 16 were 31, 27, 25, 12, 0, 13, 18, 1; 95 errors in the first 4 and 32 errors in the last 4 patterns learned. M's average errors per maze in combining were 15.9, or 20 percent of the part method totals. Time and trials were relatively larger in amount and number: 540 seconds and 3.9 trials, or 56 and 53 percent of the part method totals exclusive of the first and last three trials in initial part learning.

Subject C's performance in combining the parts was (as in the earlier investigation) highly erratic. Her error records show 6, 26, 35, 236, 39, 108, 16, 13 for the same mazes as

given above for subject M. The error average is 59.8; higher than was the case with either W or M, but almost entirely because of the presence of the extreme values 108 and 236. C's error median is 30.5 and W's 51.5, and except for 108 and 236 C's worst record is 39. For time and trials C required on the average 931 seconds and 5.6 trials for integration. The percentages of the part method totals taken for combining are 46, 62, and 60 percent for errors, time, and trials.

M's marked superiority to W in combining may be due to slower and more cautious initial learning, or (and) to a retrospectively reported habit of paying special attention to the end and beginning of each part during 8 unit learning. As we have seen, C's method of tracing did not lend itself to verbalization. Perhaps because of exceptionally vivid imagery, she seemed to rely upon motor patterns or motor to visual translations of the lay-out of each unit, often changing the direction taken around the central block (within a given unit) from trial to trial in a manner likely to disorient one relying upon counting. Apparently her method usually favored subsequent integration of parts into a whole, but lacking the stabilizing influence of a sequence of numbers, orientation once lost was difficult to regain.

(3) *Practice*

(*a*) *Whole Procedure.*—Except for one 'reversal' with subject C on maze 11, the 32 unit mazes learned by the whole method show a consistent decrease in errors, time, and trials for each of the 3 subjects. C's reversal with maze 11, however, was sufficiently striking to distort the composite pattern slightly. The averages for all 3 subjects are: errors, 312, 222, 219, 153; time, 1831, 1414, 1560, 990; trials, 11.8, 10.5, 11.2, 7.5 for mazes 1–4, 5–8, 9–12, 13–16. As noted previously, the most striking practice effect is from maze 1 to maze 2, but this does not markedly affect the values for the whole method, as W alone used whole procedure with maze 1.

(*b*) *Initial Part Learning.*—Here we need only consider errors, since the trends are similar for all 3 measures. Rather marked individual differences appear.

Subject W showed considerable practice effect. For mazes 1–8 she made an average of 5.9 errors in each 8 unit part, for mazes 9–16, only 3.0 errors. Taking the patterns by fours, however, there is a slight increase from mazes 1–4 to 5–8 and a continuous drop thereafter. The averages in order are: 5.2, 6.3, 4.2, 1.7 errors. In this connection it is interesting to note that with 8 unit parts of the 16 unit mazes (D 31–38) learned 6 months previously (2) W's average errors were 3.5, a value only reached (and surpassed) in the present experiment with mazes 13–16.

In contrast to W, subject M (if maze 1 is excluded) improved not at all in initial part learning. For the 8 unit parts of maze 1 M made an average of 29.0 errors per part, but for maze 4 only 12.0 errors per part, and for each successive pair of the other 6 mazes learned by the part method 12.8, 13.9, and 14.4 errors. With mazes D 31–38 in the earlier experiment M's average errors for initial part learning (8 unit parts) were 12.1. The difference between the values in the present experiment (excepting between maze 1 and the rest of the group) are small, but in view of M's remarkable consistency and resistance to 'chance' disturbances, probably signify a genuine retrogression. The cause of this retrogression is almost certainly increasing concentration upon the beginning and end of each part, as M noted the aid which this device gave in the act of integration.

C also showed a marked practice effect from maze 1 to maze 4, both of which were learned by the part method. Her average errors for the parts of maze 1 were 30.9, maze 4, 19.8; mazes 5–8, 9–12, 13–16; 13.4, 18.5, 12.5 errors. All except maze 1 show a decided improvement over C's error record (26.2) for the 8 unit patterns in the earlier experiment (D 31–38).

(*c*) *Combining the Parts.*—Subject W did not improve in ability to combine in parts. Subject M improved markedly, even though his skill at the beginning of the experiment was much greater than that of W. He made an average of 24 errors per maze in combining mazes 1–8, and 8 errors per maze in combining mazes 9–16. C's results are so variable

that nothing can be determined about the influence of practice. Detailed results for all three subjects have been given previously in discussing the relation between combining the parts and amount of material.

In general, the evidence indicates a practice effect for all 3 subjects with whole procedure, for W and C in initial part learning (large for W and slight for C), and for M in combining the parts. No improvement is shown by M in initial part learning (except from maze 1 to subsequent mazes) or by W and C in combining the parts. Variability from maze to maze makes C results difficult to interpret, but the difference between the influence of practice with M and W is almost certainly a function of the method of learning. W gave major attention to the task in hand, and so improved in 8 unit learning at the expense of later integration. M gave special heed to the end of one part and the beginning of the next. He therefore improved only in ability to combine the parts. It should be noted that this difference in mode of attack is an aspect of the attempt to adjust to the more difficult task presented by the 32 unit mazes: it appears only to a slight degree in the earlier experiment. It would be of interest to carry the investigation to a point where each subject approached maximum improvement in his or her chosen field of concentration, and note whether and with what result attention would be directed to the previously slighted phase of the task.

(4) *Regional Errors in Combining the Parts.*—Subject C showed a curious pattern of regional errors in the act of integration, but one whose significance I cannot interpret. For parts I, II, III, and IV her total errors on the first combining trial were 36, 64, 30, 27; on all combining trials, 70, 227, 82, 50. For the first combining trial the distribution of errors among the parts is similar to that given previously for initial part learning (see amount of material), but for all combining trials the large number of errors in part II is difficult to explain. The number (227) is too much greater than the values for the other parts to be attributed to chance, and it is not a function of the size of pattern, since it does not appear even for subject C with the whole method. C's aver-

age errors for whole learning are 41, 62, 74, and 67 for each successive 8 units of the 32 unit patterns. It can only be due to some peculiarity of her method of integration.

For M and W the most striking feature of the data is the small number of errors in part IV. For the first combining trial the total errors for parts I, II, III, IV are: W, 51, 55, 64, 17; M, 33, 24, 16, 1; total combining trials: W, 91, 133, 131, 38; M, 45, 43, 26, 1. Here the choice of interpretations is between 'recency' and 'retroactive inhibition.' For purposes of comparison it might be of interest to have the remainder of the data on regional errors with the whole method. For each of the 4 successive 8 unit sections of the whole patterns W made an average of 25, 70, 63, and 53 errors; M, 50, 56, 55, and 60 errors.

(5) *Individual Differences.*—Since individual differences have been treated in some detail throughout the previous discussion, at this point we need only note that significant individual differences appear only with the part method. W is superior to the others in initial part learning, M is superior in combining the parts.

III. Summary

For 3 experienced subjects learning spider mazes first trial differences between the relative economy of whole (32 unit) and 4 part learning were slight. For later trials part learning was in general strikingly superior to whole learning in errors and slightly superior in time. Trials showed a small but unreliable advantage for whole procedure. The chief factor of advantage for the part method is disproportionate increase in difficulty of 32 unit patterns over 8 unit patterns, and the chief 'factor of disadvantage' the difficulty of combining the parts. Thus the similarity between the totals for the 2 adults by the part method was due to the fact that while one subject expended but little effort in 8 unit (part) learning and had considerable subsequent difficulty with integration, the other adult associated a moderate increase in difficulty from 8 to 32 unit learning with few errors in the combining act. The 3 subjects showed only minor differences for whole procedure,

but one subject (a 10 year old child) gave a somewhat erratic performance and was in general inferior to the adults for both phases of part procedure, and thus showed only a relatively small advantage for the part method in errors and an advantage for the whole method in both time and trials.

Practice had little or no influence upon the relative economy of part and whole procedure, within the limits of the experiment.

(Manuscript received December 20, 1937)

References

1. Cook, T. W., Whole versus part learning the spider maze, *J. Exper. Psychol.*, 1937, **20**, 477–494.
2. Cook, T. W., Factors in whole and part learning a visually perceived maze, *J. Genet. Psychol.*, 1936, **49**, 3–32.

STUDIES IN THERMAL SENSITIVITY: 5. THE REACTIONS OF UNTRAINED SUBJECTS TO SIMULTANEOUS WARM + COLD STIMULATION

BY WILLIAM LEROY JENKINS

Lehigh University

The first systematic experiments with warm + cold synthesis were performed by Cutolo.[1] He tried both punctiform and areal stimulation upon three introspectively-trained observers. For areal stimulation, he used a grill consisting of eight glass tubes having an outside diameter of 6.5 or 7 mm. Warm water at 43–45° C. circulated through the even numbered tubes; cold water at 5–9° C. through the odd numbered tubes. With punctiform stimulation, all three observers reported 'heat' in the majority of instances. For the grill experiments, exact figures are not given, but they were evidently strongly positive because Cutolo says: ". . . even the untrained observer can easily convince himself that heat derives from warm and cold stimulation."

Alston was interested in the problem of how far punctiform warm and cold stimuli could be separated and still produce a fusion in heat. His stimulator temperatures were 43–44° C. and 5–10° C. The three introspectively-trained observers reported 'heat' in considerably more than half the cases, even with the stimulators separated by as much as 10–15 cm.[2]

Hahn, apparently in ignorance of the foregoing studies, made some tests with a double spiral of copper tubing using temperatures of 34–38° C. and 18–25° C.[3] The arm was placed on this stimulator at room temperature, and allowed

[1] Cutolo, F., A preliminary study of the psychology of heat, *Amer. J. Psychol.*, 1918, **29**, 442–448.

[2] Alston, J. H., The spatial condition of the fusion of warmth and cold in heat, *Amer. J. Psychol.*, 1920, **31**, 303–312.

[3] Hahn, H., Die Reize und die Reizbedingungen des Temperatursinnes. II. Die Reizbedingungen des Temperatursinnes, *Arch. f. d. ges. Physiol.*, 1927, **217**, 66–67.

to remain until all temperature sensations had disappeared; then cold water and warm water were allowed to flow through the tubes. His results are interesting. When both water valves were opened at the same time, conflicting warm and cold sensations occurred, with the cold eventually dominating. When the cold water was allowed to run for about four seconds, then the warm added, the cold sensation disappeared and the warm dominated. But when the warm water ran for about four seconds before the cold was added, with the addition of cold: ". . . . so verschwand schnell die Warmempfindung. Es trat aber nunmehr keine Kälteempfindung an ihre Stelle, sondern eine neue *viel intensivere* Empfindungsqualität, die sich qualitativ auf das Charakteristischte unterschied. Für diese Empfindung drängte sich uns und allen unseren Versuchspersonen *ganz spontan* die Bezeichnung 'Hitze' auf." [4]. This differed in no respect from the normal heat sensation, except that the frequently-found burning sensation was not present. Hahn makes no statement as to the number or training of his observers.

Burnett and Dallenbach, using an improved type of grill, obtained positive results with three introspectively-trained observers.[5] Their grill was made of copper tubing with 3 mm. outside diameter, having a total stimulator surface 8 × 8 cm. Temperatures of 2°, 10°, 20° and 30° C. were used five times each in combination with 34°, 37°, 40° and 43° C. Heat was reported 50, 70 and 32 times respectively out of 80 trials.

In a later study, Burnett and Dallenbach derived a formula for the intensity of synthetic heat, using one grill for each arm in the method of paired comparisons.[6] Gritman and Dallenbach later verified this formula with a carefully-balanced series of temperature combinations.[7] The formula: $(32 - C) + 3(W - 33)$—indicates that the warm tempera-

[4] *Op. cit.*, 67.

[5] Burnett, N. C. and Dallenbach, K. M., The experience of heat, *Amer. J. Psychol.*, 1927, **38**, 418–431.

[6] Burnett, N. C. and Dallenbach, K. M., Heat intensity, *Amer. J. Psychol.*, 1928, **40**, 484–494.

[7] Gritman, W. B. and Dallenbach, K. M., The formula for the intensive gradation of heat, *Amer. J. Psychol.*, 1929, **41**, 460–464.

ture plays about three times as strong a role in determining the intensity of synthetic heat as does the cold.

Ferrall and Dallenbach, in connection with their study of burning heat, made some incidental experiments in the synthesis of heat with two trained observers and one novice.[8] Because of their particular relation to the results of the current study, these experiments will be considered in greater detail later.

Sullivan and Verda stimulated individual warm and cold 'spots' with air jets to show that the 'pressury' element in heat reported by earlier investigators was an artifact of the method of stimulation.[9] Temperatures were 38–40° C. and 17–19° C. The four introspectively-trained observers reported 'heat' a total of 132 times out of 137 trials, even with spatial separations of as much as 14 cm.

Considering the weight of all this evidence, it becomes apparent that introspectively-trained observers have consistently reported 'heat' when stimulated simultaneously with warm + cold. The reactions of *totally-untrained* subjects to such compound stimulation do not seem to have been systematically investigated. The difficulties of working with untrained subjects, particularly with thermal phenomena, must be clearly recognized. Yet the following assumption appears to be sound: If these subjects report 'hot' when stimulated with an appropriate single temperature, they should also report 'hot' with warm + cold stimulation—provided that the synthetic experience is identical with the natural one.

Apparatus and Procedure

Four grills were constructed for this work. Grill No. 1 embodied two features not found in the apparatus used by earlier investigators:

1. It was made of rectangular bars, instead of round tubes, so that a flat surface was presented to the skin. With tubes, the amount of grill surface actually in contact varies with the pressure of the arm.

2. It was adjustable. One set of bars (warm) was fixed in position. The alternate set could be dropped about 5 mm below the surface, or raised by a cam action to the

[8] Ferrall, S. C. and Dallenbach, K. M., The analysis and synthesis of burning heat, *Amer. J. Psychol.*, 1930, **42**, 75–77.

[9] Sullivan, A. H. and Verda, D. J., The experiential fusion of warmth and cold in heat, *J. Exper. Psychol.*, 1930, **13**, 208–216.

exact level of the fixed bars. This permitted starting with a single temperature and adding a second; or starting with the compound and subtracting one component.

Grills 2, 3 and 4 were also made with flat bars, but did not have this adjustable feature, and were designed merely to provide some crude checks upon the results with No. 1. In grills 1 and 4, the water flowed through holes drilled lengthwise of the bars; in grills 2 and 3, the bars were solid and received their temperature by conduction from larger tubes at the sides. The chief dimensions of the four grills are shown in Table 1.

TABLE 1

CHIEF DIMENSIONS OF GRILLS

	Identifying Number			
	1	2	3	4
Warm bar width, mm	3	3	1½	4½
Cold bar width, mm	3	1½	1½	4½
Air space between bars, mm	1	1	1	2
Number of warm bars	6	6	8	6
Number of cold bars	5	5	7	6
Grill width, mm	42	30	33	75
Grill length, mm	47	25	25	63

The width of the bars in Grill No. 1 is approximately the same as the diameter of the tubes used by Burnett and Dallenbach. The width of the bars in Grill No. 4 approaches the diameter of the tubes used by Cutolo. Grills 2 and 3 are unlike any used by earlier workers. It will be noted that in Grill No. 2 one set of bars (warm) was twice the width of the other set (cold).

The temperatures of the water pumped through the two parts of the grills were controlled by two mixing valves, similar in principle to the mixing valves used in a shower-bath. Tests showed that the difference between the inlet and outlet temperatures for the warm water in Grill No. 1 was about 2° C., regardless of the cold temperature employed. In practice, only the inlet temperatures were taken.[10] Room temperature was maintained at 21–23° C.

PRELIMINARY EXPERIMENTS

By way of general orientation, 157 members of a class in elementary psychology were used as subjects. None of these individuals had received any training in introspection, although all of them had performed a laboratory experiment in 'mapping' of warm and cold sensitivity and had been exposed to a statement of the Alrutz theory. Grill No. 1 was used throughout. Each subject was stimulated once with a single warm temperature, and then immediately afterwards with a compound of the same warm temperature with cold. Four combinations were used: 38° with 15° C.; 40° with 15° C.; 44° with 15° C.; 38° with 21° C. Three-fourths of the subjects reported both experiences upon a form which suggested a series of thermal terms: very cold,

[10] Outlet temperatures have generally been taken by previous workers. Inlet temperatures were used here as a matter of convenience. Probably the temperature of the grill surface itself lies somewhere between the two. For more critical work, the surface temperatures at various parts of the grill should be measured directly by thermocouple-galvanometer technique.

cold, etc. through neutral to very warm and hot. The balance were asked merely to describe their experiences as accurately as possible.

The results were quite unexpected. 'Hot' was reported only 16 times with the compound stimulation. In 6 of these instances, the subject had also called the simple stimulation 'hot,' which left only 10 clear reports attributable directly to compound stimulation. In some 48 percent of the cases, the subject indicated that the intensity of the warmth had been decreased or obliterated by the addition of cold, and only 12 percent reported a definite increase.[11]

Principal Experiments

For the main study, arrangements were made whereby 126 members of the elementary class reported in groups of about ten for a two-hour period. To economize on time, one setting of temperatures was made on the grills, and the members of the group came in one after another for the test. Then another setting was made, and the process repeated. This permitted covering about 20 combinations of temperatures in the two-hour period. The inlet temperatures were 36°, 38°, 40°, 42°, 44°, 46°, 48° and 50° C., compounded with 10°, 15°, 20° and 25° C. Since this involved 32 possible combinations, all of them could not be used with each subject; but a representative assortment was used with each group. The various temperature combinations were presented in an irregular order to prevent the subjects from developing a set for a particular type or sequence of reports.

76 subjects were tested with Grill No. 1, which permitted a direct comparison between simple and compound stimulation. Three schemes of stimulation were used:

101—simple warm first for 5 seconds; arm then lifted off grill momentarily while the cold bars were brought up; then arm replaced on grill for 5 seconds more.

103—simple warm first for 5 seconds; then with arm remaining in position, the cold bars were brought up, and the arm allowed to rest on the compound for 5 seconds.

105—compound stimulation first for 5 seconds; then the cold bars were dropped, and the arm allowed to rest upon the simple warm grill for 5 seconds.

29 subjects were stimulated with grills 2 and 3, and 21 subjects with grills 2 and 4. The subject placed his arm upon one grill for 5 seconds; then transferred it to the other for 5 seconds.[12] Direct comparison with simple warm stimulation was not possible in these cases. During each series, however, the subjects were stimulated with various warm temperatures compounded with cold, and also with the same inlet warm temperatures compounded with 34° C. An indirect comparison was thus possible.

There are obvious artifacts in this crude method of comparison, which rests upon the assumption that the surface temperature of the warm bars is not affected by the temperature of the cold bars. This is probably not true, especially with extreme degrees of cold. In spite of the 1 mm air gap between the bars, there would be some interchange. This applies particularly to grills 2 and 3, which received their temperatures by conduction. In Grill No. 1, where direct simple vs. compound comparison could be made, this artifact was avoided.

---

[11] Needless to say, throughout all the experiments, the subjects were not informed of the nature and purpose of the technique.

[12] Except for the tests with grills 2 and 4, the grills were concealed by a hood. With grills 2 and 4, they were deliberately exposed to see if this would affect the reports.

All stimulations were made upon the volar surface of the left forearm slightly above the wrist. All reports were made upon the same form. The instructions printed on the form were:

"Try to report accurately all thermal sensations in each experience, using the following symbols:

V.C. —very cold	B.C.—barely cool	WARM—warm
COLD—cold	O —neutral	V.W. —very warm
COOL—cool	S.W.—slightly warm	HOT —hot

"Add under 'Remarks' any comments that you think will help to clarify your report, including record of non-thermal sensations (pain, etc.)."

RESULTS

In the following analysis of results, 'simple stimulation' means the use of a single temperature (36°, 38°, 40°, 42°, 44°, 46°, 48°, or 50° C. at the inlet). With Grill No. 1 this occurred when only the fixed bars were in contact with the skin. With grills 2, 3 and 4, the same situation was approximated by having alternate bars at skin neutral. 'Compound stimulation' means that one set of bars was warm, and the other cold (10°, 15°, 20°, or 25° C.). As the results were not significantly different with these four low temperatures, they are not analyzed separately in the tables.

By a 'clear chance' for a report of hot with compound stimulation is meant a case where 'hot' was not reported with the just preceding or following simple stimulation in case of Grill No. 1; or with the same inlet temperature combined with skin neutral in grills 2, 3 and 4. Similarly, a 'clear report of hot' with compound stimulation indicates that hot was not reported with the corresponding simple stimulation.

Table 2 shows the gross results for all four grills. 'Hot' was not reported in any significant percentage of cases with inlet temperatures lower than 44° C. Above this level, the percentage is actually higher with simple than with compound stimulation.

Table 3 shows only *clear* reports of hot with compound stimulation in percentage of total chances for such clear reports. Again, the figures do not reach any significant level with inlet temperatures below 44° C., and exceed 20 percent in only one instance (48° C. with Grill No. 1). Practically all of these cases with 48° C. occurred when the compound stimu-

TABLE 2

PERCENTAGES—TOTAL REPORTS OF 'HOT'

Grill	Inlet Temperatures							
	36°	38°	40°	42°	44°	46°	48°	50°
No. 1 Simple	0.5	1.2	3.9	7.1	16.3	30.0	70.0	94.1
No. 1 Compound	0.5	2.0	4.6	7.1	10.2	13.4	37.0	74.7
No. 2 Simple	0.0	0.0	2.0	8.0	18.0	24.0	66.0	100.0
No. 2 Compound	0.0	2.1	6.0	8.8	14.9	17.6	56.3	83.3
No. 3 Simple	0.0	0.0	0.0	2.4	2.4	14.3	45.3	60.0
No. 3 Compound	2.2	3.2	1.8	2.7	9.1	11.6	12.5	26.0
No. 4 Simple	0.0	0.0	0.0	0.0	0.0	28.5	76.3	—
No. 4 Compound	14.3	7.9	7.5	9.5	12.7	18.9	27.2	—

TABLE 3

PERCENTAGES—CLEAR REPORTS OF 'HOT' WITH COMPOUND STIMULATION

Grill	Inlet Temperatures							
	36°	38°	40°	42°	44°	46°	48°	50°
No. 1	0.5	2.1	4.4	6.0	11.7	13.7	34.5	—
No. 2	0.0	2.3	6.0	8.0	11.5	10.0	16.7	—
No. 3	0.0	2.4	0.0	0.0	2.8	20.0	5.6	—
No. 4	14.3	7.9	7.6	9.5	12.7	17.4	—	—

lation was given first. Since 48° C. simple stimulation was called 'hot' 70 percent of the time, the apparently 'clear' reports with compound stimulation must be viewed with some suspicion.

Table 4 gives perhaps the clearest picture of the situation.

TABLE 4

NUMBER OF SUBJECTS GIVING CLEAR REPORTS OF HOT WITH COMPOUND STIMULATION

Times reported	With Grills				
	No. 1	No. 2	No. 3	No. 4	All
Never	37	32	25	11	61
Less than 10% of clear chances	23	7	3	5	41
10%–24% of clear chances	10	8	1	2	15
25%–49% of clear chances	6	3	0	2	8
50% or over	0	0	0	1	1
Total subjects	76	50	29	21	126

Nearly half of the 126 subjects never gave a clear report of hot with any compound stimulation. Most of the remaining individuals reported it only sporadically. Less than a dozen subjects gave clear 'hot' reports more than 25 percent of the time. The best record is that of a single individual who reported hot 11 out of 13 clear chances with Grill No. 4 (but only 3 out of 15 times with Grill No. 2). With Grill No. 1, the best individual record was 7 reports of hot out of 19 clear chances.

Table 5, based upon Grill No. 1 only, shows the comparative intensity of warmth with simple and compound stimula-

TABLE 5

INTENSITY OF WARMTH IN COMPOUND STIMULATION COMPARED WITH SIMPLE STIMULATION

Order		Inlet Temperatures							
		36	38	40	42	44	46	48	50
Simple First:	% Stronger	3	3	6	3	3	2	6	0
	% Same	22	21	11	6	5	9	20	82
	% Weaker	75	76	83	91	92	89	74	18
Compound First:	% Stronger	8	22	38	39	46	64	19	6
	% Same	38	10	15	10	15	5	27	62
	% Weaker	54	68	47	51	39	31	54	32
* Average:	% Stronger	6	13	22	21	25	33	13	3
	% Same	29	15	13	8	9	7	23	72
	% Weaker	65	72	65	71	66	60	64	25

* To compensate for influence of order.

tion. When the simple stimulation was given first, the effect of adding cold was strongly in the direction of masking, with only a few scattering reports of accentuation. This might be largely an artifact of the order of stimulation. However, even when the compound stimulation was given first, the succeeding simple warmth was reported more intense in a surprisingly large percentage of the cases. An average of the two sets of figures should neutralize any influence of the order of presentation. The averaged percentages show clearly a preponderance of masking, occurring rather uniformly in 60

to 72 percent of the cases. The minority accentuation effect rises to a peak at 44–46° C., but never exceeds 33 percent.

Discussion of Results

It is obvious that the foregoing results are not at all in agreement with those obtained with introspectively-trained observers. These untrained subjects did not consistently report hot when stimulated with warm + cold. In fact,the addition of cold seemed to decrease, rather than to accentuate, the intensity of the warmth experienced with comparable simple stimulation.

Two possible explanations present themselves:

1. *Differences in Technique.*—But the fact is that the earlier workers obtained consistent reports of 'heat' from trained observers with techniques varying from the simultaneous stimulation of single warm and cold 'spots' . . . to the use of grills with tubes 7 mm in diameter. If it were simply a matter of technique, it seems likely that some *one* of our four grills should have given positive results.

Furthermore, a search of the literature discloses a single case of a totally-naive subject. This is in the study by Ferrall and Dallenbach.[13] Three observers were used: D, trained in introspection and with previous experience in grill experiments; J, trained in introspection but with no previous experience in cutaneous experiments; A, 'altogether untrained in introspection.'[14] D reported 'heat' in 15 out of 18 compound stimulations; J in 12 out of 16; but A (the novice) only 3 times out of 21 compound stimulations![15] This percentage of reports (14.3 percent) is low enough to be consistent with some obtained in the present study. This is added reason to suspect that the discrepancy does not lie solely in differences of technique.

2. *Differences in Training.*—There are at least three ways in which the variant results could be explained in terms of differences in training:

[13] Ferrall, S. C. and Dallenbach, K. M., The analysis and synthesis of burning heat, *Amer. J. Psychol.*, 1930, **42**, 72–82.

[14] *Op. cit.*, 74.

[15] *Op cit.*, 76.

(*a*) With compound stimulation, the untrained subjects actually have an experience identical with the normal 'hot,' but fail to call it by its proper name. However, if the natural and synthetic experiences are truly identical, this suggestion is somewhat fantastic. Furthermore, the subjects refused to call the compound experience 'hot' in the face of direct suggestion. On the final stimulation of each experimental period, the subjects were told: "This may be very hot. Take your arm off the stimulator if it becomes too severe." Results were flatly negative.

(*b*) The 'heat' reported by trained observers with compound stimulation is merely an illusion. (Just as an illusion of movement can be produced stereoscopically by presenting a vertical line to one eye and a horizontal line to the other.) However, it would ill become an author, who is himself untrained to introspection, thus gratuitously to indict the validity of the trained observers' descriptions. Such an explanation (if it is to be made at all) should come from the trained observers themselves.

(*c*) The simplest explanation seems to be this: The trained observer's 'heat' is a *different experience* from the untrained subject's 'hot.' From simple stimulation, the introspectively-trained observer derives an experience which he calls 'heat.' Because of his specialized training, he is able to examine the experience carefully and pronounce it qualitatively different from the experience he calls 'warm.' Then, with certain compound stimulations, he receives a perception which is identical in its chief characteristics with the natural 'heat'; so he calls it by the same name.

Untrained subjects, on the other hand, lack the advantage of this specialized training. With simple stimulation, they have an experience which they call 'hot,' which seems, in the main, to be merely a higher degree of the experience they call 'warm.' With compound stimulation, they do not (except rarely) have the identical experience—if anything, the intensity of the warmth seems to be reduced—so they refuse to use the term 'hot' in describing it.

The outcome of this last hypothesis is obvious. For some years, the Alrutz theory of 'heat' has been widely accepted.

One of the main supports of the theory has been the series of synthetic experiments . . . with trained observers . . . which have been taken as an indication that 'heat' is dependent physiologically upon the simultaneous stimulation of warm and cold receptors. It seems to have been tacitly assumed that the same synthetic results could be obtained with untrained individuals, and therefore that the Alrutz theory holds for the common experience of 'hot.'

In the experiments just reported, positive synthetic results could not be obtained with untrained subjects. This does not prove that the Alrutz theory does not hold for the common experience. The methods of synthesis may have been faulty. Someone else may develop a better technique and secure positive effects with untrained subjects. But until such positive results are actually obtained, the synthetic experiments with trained observers should not be quoted in support of the Alrutz theory as applied to the common experience of 'hot.'

Summary

126 untrained subjects were tested with warm + cold stimulation, using, in all, four different types of grills. The temperatures were 36°, 38°, 40°, 42°, 44°, 46°, 48° and 50° C., compounded with 10°, 15°, 20° and 25° C. Direct or indirect comparison with the corresponding simple stimulation was made in each case. 'Hot' was *not* reported in response to the compound stimulation, except in a scattering minority of cases. Generally, the intensity of the warmth reported with compound stimulation was *less* than that with corresponding simple stimulation.

Earlier investigators, using introspectively-trained observers, found that 'heat' was reported in response to such compound stimulation in an overwhelming majority of cases. In explanation of the discrepancy, it is suggested that the 'heat' of the trained observer is a different experience from the common experience 'hot.' Synthetic experiments with trained observers, therefore, should not be cited as evidence that the Alrutz theory (simultaneous stimulation of warm and cold receptors) applies to the common experience of 'hot.'

(Manuscript received December 31, 1937)

NOTE ON THE RELIABILITY AND THE VALIDITY OF THE GROUP JUDGMENT

BY MALCOLM G. PRESTON

University of Pennsylvania

Gordon [1] has recently replied to Stroop's criticism of her interpretation of the results of an experiment done some years ago on the group judgment. In the original experiment [2] Gordon required a large number of subjects to order a series of weights. A coefficient of correlation (by the rank difference method) was calculated between the true order and the obtained order in the case of each subject and the arithmetic mean was obtained from the distribution of coefficients. This value amounted to + .41. The subjects were then divided into groups of equal size. By simple averaging of the reports of the individual subjects in each group, as many orders were obtained as there were groups. Each of these group orders was correlated with the true order and the mean of the coefficients computed. This value turned out to be + .68. Amalgamation of the original orders into groups comprising more and more data resulted in a steady increase of the mean coefficient of correlation. On the basis of this observation Gordon concluded that the judgment of the group is better than the judgment of the average member of the group.

Stroop [3] maintained that although this conclusion had received some acceptance, yet it was unsound. He rested his argument upon the fact that Gordon's amalgamation technique served to fulfill the requirements of the Spearman-Brown Prophecy Formula. The observed increase in the

[1] GORDON, K., Further observations on group judgments of lifted weights, *J. Psychol.*, 1935–36, 1, 105–115.

[2] GORDON, K., Group judgments in the field of lifted weights, *J. Exper. Psychol.*, 1924, 3, 398–400.

[3] STROOP, J. R., Is the judgment of the group better than that of the average member of the group, *J. Exper. Psychol.* 1932, 15, 550–562.

mean value followed mathematically upon an increase in the number of cases and, as Stroop pointed out, could be shown as well in a card sorting demonstration as in the experiment which Gordon reported. To bring point to this assertion Stroop reported the order in which cards fell in a series of 100 sortings. The mean coefficient of correlation was valued at − .034. Amalgamation of Stroop's data following Gordon's method produced an increase in the absolute value of the coefficient of correlation, which increase depended upon the number of individual orders directly affecting the coefficient. He concluded that "the increase in validity obtained by grouping data according to Gordon's plan is not due to the fact that the judgments were made by different persons but is due to the mathematical principle involved in this method of combining data."

T. L. Kelley has made an analytic study of this problem [4] and Stroop's card-sorting experiment has furnished empirical verification to his findings. Kelley showed, among other things, that the expected mean coefficient could be calculated by means of the Spearman-Brown formula. He compared the obtained values published by Gordon with his expected values and observed that the maximum departure from expectation was less than 2 sigma. Gordon's values, therefore, appeared to approximate those magnitudes which were expected on the basis of the operation of the Spearman-Brown formula. Kelley concluded that "we may state Gordon's conclusion explicitly and say that the increase in reliability is that forecast by the Spearman-Brown formula."

Gordon's reply to Stroop's argument rests largely on the results of a card sorting experiment, similar to Stroop's, in which she failed to note any consistent increase in the mean coefficient as the number of orders affecting the individual coefficient were increased. Following this demonstration, Gordon publishes an additional experiment using the technique under discussion. A paper by Bruce [5] has also appeared in which the same technique has been applied.

[4] KELLEY, T. L., The applicability of the Spearman-Brown formula for the measurement of reliability, *J. Educ. Psychol.*, 1925, 16, 300–303.

[5] BRUCE, R. S., Group judgments in the fields of lifted weights and visual discrimination, *J. Psychol.*, 1935–36, 1, 117–121.

The conflict between Gordon's card-sorting results and those reported by Stroop is, however, more apparent than real. Stroop reports a mean coefficient of $-.034$ which increases in absolute value as the individual groups comprise more and more data. This is what would be expected in view of the Spearman-Brown Prophecy Formula. Gordon reports a mean coefficient of zero with no consistent increase in either direction as the number of orders to the group is increased. It is on the basis of this observation that she questions the substance of the criticisms applied to her work. But this lack of increase is to be expected in view of the formula itself. The formula gives for the expected r the value: $\frac{r_{11}}{1 + (n - 1)r_{11}}$. Inspection of the formula indicates that a value other than zero for the predicted coefficient depends upon a value other than zero appearing in the numerator of the fraction composing the formula. In the event that the observed value is zero, the expected value must be zero.

Because additional experiments have appeared since Stroop published his criticism and because Stroop did not clearly distinguish between what was to be expected in the event the mean coefficient did not equal zero from what was to be expected in the event the mean coefficient equalled zero, verification is reported here whose purpose is to show that if the mean of the coefficients computed on the individual orders is equal to zero, no consistent change will appear in the mean of the coefficients based on average orders. If on the other hand, the mean of the coefficients computed on the individual orders is greater or less than zero, then amalgamation of the orders following Gordon's method will result in an increase in the mean value of the coefficients obtained thereby. Since the demonstration about to be reported was done with cards it follows that the variations being studied are a result of statistical rather than psychological factors.

Methods and Results

A pack of conventional playing cards, 52 in number, with 13 to each suit, was sorted 25 times. Before each sorting, the

cards were ruffed five times and between each ruffing they were shuffled five times. The order in which the cards composing each suit appeared was kept. Aces were written as ones, jacks as elevens, queens as twelves and kings as thirteens. Since there are four suits, the twenty-five sortings produced 100 chance orders. The 100 orders are not produced here because they differ in no important respect from orders already published by Gordon.

Each obtained order was correlated by the rank difference method with the true order 1–13. When this had been finished and the mean coefficient obtained it was found to be − .014. We had established as an arbitrary criterion of mean zero correlation a value between − .004 and .004. Consequently one order was eliminated and an additional sorting made to produce an order to replace the extreme order eliminated. The final data produced a mean coefficient of − .002, which of course met the criterion previously established.

Table 1 shows the results of this sorting. In column 1 appear the step intervals into which the 100 coefficients of

TABLE 1

GIVING FREQUENCY DISTRIBUTION OF ALL COEFFICIENTS OF CORRELATION OBTAINED ON SINGLE ORDERS AND ON AVERAGED ORDERS

Value	1 Order per Coefficient	4 Orders per Coefficient	10 Orders per Coefficient	20 Orders per Coefficient
850–949				
750–849				
650–749	2			
550–649	1			
450–549	4		1	1
350–449	5	3		
250–349	7	3	3	1
150–249	11	4		
50–149	15	3	1	1
− 50–49	12	1		1
−150–(−51)	12	2	2	
−250–(−151)	8	3	1	
−350–(−251)	11	3		
−450–(−351)	5	3	1	
−550–(−451)	4		1	
−650–(−551)	2			1
−750–(−651)	1			
−850–(−751)				
Mean	−.002	+.012	+.020	+.060
ave.	.03			

correlation calculated by Gordon's method were classified. Column 2 gives the frequency for each step interval. Column 3 contains 25 coefficients each based upon the average of 4 orders. Column 4 contains 10 coefficients each based on the average of 10 orders. Column 5 contains 5 coefficients each based on the average of 20 orders. The means of the columns in order are: − .002, + .012, + .020 and + .060. The latter three averages lie within the limits established by the standard error of the mean of the original distribution (which appears at the foot of column 2) and confirm what Gordon has previously shown, *i.e.* when the mean coefficient is zero, amalgamation of the data by averaging orders does not produce any significant increase in the mean coefficient of correlation.

Following the construction of this table, the orders were separated into three groups, those in which the coefficient was zero, those in which it was negative and those in which it was positive in sign. It was observed that 4 of the coefficients were zero and that 48 each were positive and negative. The zero coefficients were discarded and separate treatment was given to the positive and negative coefficients. Distributions in each case were obtained for coefficients based upon 1, 4, 8, 16 and 48 orders.

Table 2 shows the result of averaging negative coefficients. A column is devoted to each of the distributions, and the coefficients are shown in rank order. The last value in each column is the mean of the values composing the column. Notice that these means are in order − .244, − .382, − .456, − .673 and − .855. In short, if the mean coefficient is negative in sign, averaging the orders produces means which increase in absolute value as the number of orders used in each coefficient increases. This observation confirms Stroop's observation in the case of a mean coefficient equal to − .034.

Table 3 duplicates Table 2 with the exception that the values exhibited were calculated on data comprising exclusively the positive coefficients. Notice that the means are in order .228, .401, .460, .699 and .895.

TABLE 2

GIVING RANK ORDER DISTRIBUTION OF NEGATIVE COEFFICIENTS OF CORRELATION OBTAINED ON SINGLE ORDERS AND ON AVERAGED ORDERS

1 Order per Coefficient	4 Orders per Coefficient	8 Orders per Coefficient	16 Orders per Coefficient	48 Orders per Coefficient
−.006	−.192	−.170	−.588	−.855
−.006	−.222	−.385	−.645	
−.022	−.286	−.440	−.785	
−.022	−.323	−.533		
−.050	−.336	−.518		
−.060	−.352	−.632		
−.060	−.432			
−.065	−.437			
−.071	−.460			
−.071	−.480			
−.077	−.490			
−.083	−.578			
−.088				
−.088				
−.088				
−.093				
−.093				
−.138				
−.165				
−.170				
−.182				
−.203				
−.216				
−.231				
−.238				
−.251				
−.251				
−.264				
−.274				
−.275				
−.306				
−.308				
−.308				
−.312				
−.318				
−.352				
−.374				
−.378				
−.414				
−.423				
−.458				
−.460				
−.472				
−.472				
−.526				
−.576				
−.625				
−.724				
Mean −.244	−.382	−.446	−.673	−.855

TABLE 3

GIVING RANK ORDER DISTRIBUTION OF POSITIVE COEFFICIENTS OF CORRELATION OBTAINED ON SINGLE ORDERS AND ON AVERAGED ORDERS

1 Order per Coefficient	4 Orders per Coefficient	8 Orders per Coefficient	16 Orders per Coefficient	18 Orders per Coefficient
.019	.121	.262	.515	.895
.021	.190	.329	.735	
.022	.228	.382	.848	
.027	.254	.455		
.033	.328	.521		
.048	.349	.812		
.060	.356			
.093	.416			
.094	.509			
.110	.596			
.110	.696			
.110	.773			
.115				
.115				
.121				
.122				
.126				
.129				
.131				
.143				
.148				
.166				
.170				
.170				
.176				
.187				
.192				
.209				
.226				
.232				
.238				
.264				
.306				
.308				
.312				
.312				
.330				
.356				
.368				
.405				
.416				
.428				
.500				
.500				
.510				
.525				
.590				
.667				
Mean .228	.401	.460	.699	.895

Discussion

The fact that we can predict what will happen if we amalgamate card sorting data according to Gordon's method from a knowledge of the central tendency of the original distribution of coefficients shows that the fact under discussion is undoubtedly statistical rather than psychological in character. This assertion does not, of course, deny the usefulness which psychologists may find the fact to possess. There is one aspect of the fact, however, which so far has escaped discussion, and which throws some light upon its utility. Although the Spearman-Brown formula is used as a tool for predicting the number of raters necessary to yield ratings of a given reliability, Gordon's technique, as Stroop has observed, implies its application in obtaining the highest measure of validity from a group of series of judgments. This application is supported by independent authority.[6] No one as yet, however, has considered the question of what limits exist which affect the generality of this aspect of her work. This is an important question for the social psychologist, as Goodwin Watson [7] has pointed out.

In the first place, the fact that the average coefficient is negative implies of course that the judgment of the group is *worse* than the judgment of the average member of the group, not better. If, for example, we were treating judgments of capacity, or temperament made on the basis of photographs, where experiment has shown [8] that the judgments cluster around a chance arrangement, amalgamation of a sufficiently large number of judgments following Gordon's method might well result in the group judgment being substantially worse than the judgment of the average member of the group. This would happen in the case where the median judgment showed a slight negative correlation with the true order of the subjects being judged.

[6] Kelley, T. L., *Statistical Method*, 1923, p. 199.

[7] Watson, G., Do groups think more efficiently than individuals, *J. Abnor. and Soc. Psych.* 1928–29, **23**, 328–336.

[8] Viteles, M. S. and Smith, K. R., The prediction of vocational aptitude and success from photographs, *J. Exper. Psych.*, 1932, **15**, 615–29, Esp. 622.

A second limitation on this method arises from the very nature of the concept of validity itself. The worth of a measure of the central tendency of validity depends upon two factors, first the statistical reliability of the coefficients themselves and second the confidence which we are willing to attach to the criterion. The question of the statistical reliability of the coefficients is not at issue here and may be dismissed without further discussion. The second point however deserves careful consideration. In the case of the experiments reported by Gordon and by Bruce, the measures of validity were beyond dispute because it was possible to set up a true order of events with which the Ss' judgments could be compared. But if the Ss had been judging the worth of musical compositions, or interviewing men as prospective employees or attempting to rate personality traits upon the basis of photographs, in which cases the true order is unknown or is a matter of dispute, the situation would have been qualitatively different. One might conclude from Gordon's work that in cases of this kind, the judgments of 100 interviewers if treated by her method were bound to be better than the judgments of the median interviewer of the group. We have already shown that in the presence of an acceptable criterion this conclusion does not necessarily follow. In the absence of an acceptable standard against which the original observations could be correlated, there exists a still more unsatisfactory situation. It might very well happen that the observation of the median man was positively correlated with an assumed criterion which exhibited a very restricted degree of relationship to the trait in question, and at the same time was negatively correlated with the trait itself. In this case amalgamation of the results of a number of raters would justify conclusions which were completely contradictory to the true state of affairs.

This latter possibility introduces a final question. If it is possible to produce an increase in the correlation between an assumed criterion and a series of judgments in a case where the validity is certainly not being increased, the question of what is increasing deserves to be answered. Obviously what

is being measured by the use of the coefficient of correlation is the concensus of the group in respect of the assumed criterion. In view of the fact that the existence of any correlation at all implies the existence of a certain orderliness obtaining among the distribution of judgments, and that the orderliness increases under the conditions of the technique, we can conclude that the increases in question are really increases in reliability. This follows from the fact that reliability depends upon the presence of similar orderliness among the members of separate distributions.

It appears therefore, that from the standpoint of the social psychologist the usefulness of this technique is seriously limited. Where the true order is already known the judgment of the group is more highly correlated with the true order than is the judgment of the average member of the group. But such knowledge has little practical applicability in connection with the problem of finding out the true order. Where, on the other hand, the true order is not known *a priori* the method cannot be used to learn the true order *a posteriori*.

There remains a single question as to the conditions under which it is or is not meaningful to speak of a group judgment. The increase studied by Gordon, Kelley, Stroop and Bruce are undoubtedly due to statistical factors. It seems to us, therefore, that it is improper to speak of them as a phenomenon of the psychology of the group. Perhaps our best reason for this position is the fact that in the experiments reported on this question, no possibility exists for an interaction of the viewpoints characterizing the individuals composing the group. We do regard it as a question of psychological fact whether or not judgments which are a consequence of the interaction of viewpoints are substantially better than the judgments of the average member of the group independently formulated, and we hope in the near future to contribute to this question.

(Manuscript received December 15, 1937)

EFFECT OF VISUAL AFTER-SENSATIONS UPON BRAIN POTENTIAL PATTERNING UNDER DIFFERENT DEGREES OF ATTENTION

BY LEE EDWARD TRAVIS AND MARGARET E. HALL

University of Iowa

Two previous studies have reported the after-effect of stimuli upon brain potentials. In one of them (1) a simple light, and in the other, (2) visually presented words were used as stimuli. Both studies showed that the removal of the stimulus left brain potential patterning resulting from the stimulus still functioning for some time. In neither study did the subjects report after-sensations, probably because of the brief duration of the stimuli and the mental set demanded of them by the experimental situations. It was not certain, however, that after-sensations were not present and effective.

In the present study, the effect of visual after-sensations upon the alpha brain potentials was studied under two conditions; a high degree (condition *A*) and a low degree (condition *B*) of attentiveness to after-sensation changes (positive-negative shifts). These two experimental conditions were used to isolate the factor of attention from the functional properties of the retina in regard to their respective effects upon the brain potential patterning.

Experimental Conditions

In condition *A*, the subject was requested to fixate the white light (naked frosted bulb, 1.25 foot candles, 3 meters from eyes) as long as it remained on (two durations of 10 and 5 sec.) and to signal once, by pulling a cord, as soon as the first after-sensation appeared. He was to attend closely, with eyes open, to the after-sensation and to signal each marked change observed in it. Those subjects who experienced sharp positive-negative shifts signalled immediately whenever any shifts occurred. Finally, the subject signalled the disappearance of the after-sensation. For condition *A* it was assumed that the subject would pay a high degree of attention to after-sensation changes.

In condition *B*, the same instructions were given as in condition *A*, except that the subject was requested to signal only the first appearance and the final disappearance of the after-sensation. He was to regard it passively, attending to it only sufficiently

to note its final disappearance. Shifts or changes were not to be noted or signalled. For condition *B*, it was assumed that the subject would pay a low degree of attention to after-sensation changes.

The after-sensation was, presumably, the same under both conditions, the difference in conditions bringing about a difference only in the degree of attention paid the after-sensation. Therefore, if different results were found in the two conditions, they could be attributed logically to the attentional factor.

Experimental Procedure

Each subject was given preliminary training in observing and signalling his after-sensations. A control experiment was made to determine the possible effects of signalling. The results showed that the manual reaction itself did not materially affect the alpha brain potentials.

Throughout the entire experimental routine, an attempt was made to flash the light only when the alpha rhythm was present. Occasionally, however, the stimuli were given when the alpha rhythm was absent.

The following experimental routine was followed for each subject, the two conditions being presented in *A*, *B*, *B*, *A* order to avoid unequal effects of adaptation and fatigue:

Series I:	(10 second light)	Condition *A*	2 trials
		Condition *B*	4 trials
		Condition *A*	2 trials
Series II:	(5 second light)	Condition *A*	2 trials
		Condition *B*	4 trials
		Condition *A*	2 trials
Series III:	(10 second light)	Condition *B*	2 trials
		Condition *A*	4 trials
		Condition *B*	2 trials
Series IV:	(5 second light)	Condition *B*	2 trials
		Condition *A*	4 trials
		Condition *B*	2 trials

Two subjects served in the above routine once and two others twice. The entire routine was too long to be completed in one sitting, so, in most cases, the first two series were given in one sitting and the last two in another. The brain potentials were led off from the left occipital area, one electrode being placed on the occiput and the other on the lobe of the left ear.

Results and Discussion

The Role of Attention in Brain Potential Changes During After-sensation Periods.—In Table I we note especially that for a relatively high degree of attention (condition *A*—10 mins.) the total duration of alpha waves and the mean duration of the bursts during the after-sensation periods are less

TABLE I

COMPARISON BETWEEN THE TWO CONDITIONS (HIGH AND LOW DEGREE OF ATTENTION) IN REGARD TO VARIOUS BRAIN POTENTIAL AND AFTER-SENSATION MEASURES

Condition	Total Duration of Alpha Waves During After-sensation Periods (in sec.)		Duration of After-sensations (in sec.)		Percent of After-sensation Period Alpha Waves Present	Number of Bursts During After-sensation Periods		Duration of Bursts During After-sensation Periods * (in sec.)		Length of Time between Cessation of Light and First Burst (in sec.)		Length of Time between Cessation of Light and Appearance of After-sensation (in sec.)	
	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
A–10 sec. (*N*=49)	17.6	10.9	61.1	25.9	28.8	6.1	3.7	2.5	2.5	14.8	12.4	3.6	3.0
B–10 sec. (*N*=47)	24.7	14.6	53.0	31.7	46.6	6.3	4.3	4.2	3.7	9.3	7.2	4.7	5.2
Diff.	7.1		8.1		17.8	.6		1.7		5.5		1.1	
CR	2.7		1.4		1.8	.8		6.3		2.6		1.3	
A—5 sec. (*N*=50)	14.4	9.4	40.5	19.6	35.6	4.4	2.9	2.5	2.4	12.0	10.1	3.2	2.5
B—5 sec. (*N*=43)	14.8	8.9	35.7	18.6	41.5	4.3	3.3	2.6	3.2	7.9	6.0	3.2	1.7
Diff.	.4		4.8		5.9	.1		.1		4.1			
CR	.21		1.2		.6	.1		.3		2.5			

* *A*—10 sec., *N* = 259.
B—10 sec., *N* = 294.
A— 5 sec., *N* = 214.
B— 5 sec., *N* = 184.

and the length of time for the first burst to appear after the light went off is greater than for a relatively low degree of attention (condition *B*—10 mins.). Each of these findings shows the same thing, namely, a greater effectiveness in altering brain potential patterning under condition *A* than under condition *B*. Since these two conditions differed only in the amount of attention paid the after-sensation changes it may be concluded that the attentional factor played a role over and above that of the functional properties of the retina. It is possible to measure this role fairly accurately.

(1) For total duration of alpha waves:
Minimal attention + retinal activity = 24.7 sec.
Maximal attention + retinal activity = 17.6 sec.

(2) For duration of bursts:
Minimal attention + retinal activity = 4.2 sec.
Maximal attention + retinal activity = 2.5 sec.

(3) For length of time between disappearance of light and first burst:
Minimal attention + retinal activity = 9.3 sec.
Maximal attention + retinal activity = 14.8 sec.

In (1), condition *A* was 29 percent, in (2) 40 percent, and in (3) 37 percent more effective than condition *B*. Since presumably the retinal activity was constant from the one condition to the other the difference in effectiveness must be assigned to the difference in degree of attention. Under the conditions of this experiment it is logical to conclude then that for the degree of attention existing between minimal attention in condition *B* and maximal attention in condition *A*, the attentional factor produced at least 35 percent (average of the three percentages) of the total effect.

In a study of this type the intensity of the light is probably an important consideration. The weaker the light the greater the role played by attention since retinal activity would be relatively less. Presumably one could use a light intensity value which would permit attention to be the predominant factor and another value which would make the retinal and

the attentional factor equal. In short, a graded series of light intensities might be used that would give various relationships between the effect of attention and that of the functional properties of the retina, ranging all the way from a condition in which attention was strikingly predominant to one in which it was strikingly subordinant.

In this connection, the duration of the light might be of some importance too, but in this study the two conditions were not so well differentiated for the shorter as for the longer duration.

Condition of Waves at Appearance of Light.—An attempt was made to flash the light only when there was a strong alpha rhythm present. There were times, however, when the light actually appeared during a poor, or no alpha rhythm.

TABLE II

PERCENT OF TRIALS IN WHICH ALPHA RHYTHM WAS PRESENT AT APPEARANCE OF LIGHT

Condition *A* (10 sec.)	80	*N*=50
Condition *B* (10 sec.)	83	*N*=47
Condition *A* (5 sec.)	90	*N*=50
Condition *B* (5 sec.)	93	*N*=45

Effect of Light Stimulus Upon Alpha Rhythm.—In the majority of instances the alpha waves disappeared and remained out for the duration of the light. In some instances a short burst of alpha waves appeared while the light was on. These results are given in Table III.

TABLE III

CONDITION OF ALPHA WAVES DURING LIGHT (IN PERCENT OF TRIALS)

	Alpha Rhythm Disappeared and Stayed Out	Alpha Rhythm Disappeared But a Few Waves Came in Again During Light	Alpha Rhythm Already Gone or Only Small Alpha Waves Remaining	
Condition *A* (10 sec.)	76	20	4	*N*=50
Condition *B* (10 sec.)	66	26	8	*N*=47
Condition *A* (5 sec.)	82	14	4	*N*=50
Condition *B* (5 sec.)	73	18	9	*N*=45

Relation of First Burst to Appearance of After-sensation.—In most instances the first burst of alpha waves came after

the first signal indicating the appearance of the after-sensation. In some cases, however, this chronological relationship was reversed.

TABLE IV

PERCENT OF RECORDS IN WHICH FIRST BURST PRECEDED AFTER-SENSATION

Condition *A* (10 sec.)	15	$N=46$
Condition *B* (10 sec.)	28	$N=47$
Condition *A* (5 sec.)	16	$N=49$
Condition *B* (5 sec.)	20	$N=44$

Condition of Waves During After-sensation Period.—In addition to the more specific considerations presented in Table I, relative to the condition of the brain potentials during after-sensation periods, some general evaluations are given in Table V.

TABLE V

CONDITION OF WAVES DURING AFTER-SENSATION PERIOD (IN PERCENT OF TRIALS)

	No Alpha Waves at All	No Waves, or Small, Rapid Waves with Bursts of Small, Medium, or Large Alpha Waves	Small Amplitude Alpha Waves Always Present, with Bursts of Medium or Large Alpha Waves	
Condition *A* (10 sec.)	0	96	4	$N=49$
Condition *B* (10 sec.)	4	83	13	$N=47$
Condition *A* (5 sec.)	4	94	2	$N=50$
Condition *B* (5 sec.)	2	98	0	$N=43$

Relation of After-sensation Shift-signals to Bursts.—This relationship was studied under condition *A* only, since no signals for after-sensation changes were made under condition *B*. For the 10 sec. stimulus, the signals occurred during the absence of the alpha rhythm in 90 percent of the trials; for the 5 sec. stimulus in 87 percent of the trials. Since it had been demonstrated that signalling itself had no appreciable effect upon the rhythm, it appears that the absence of the alpha waves was related to the after-sensation or attention to it. In the remaining percentage of trials, the signal came during or at the beginning or end of a burst. The alpha rhythm had usually reappeared a short time before the final signal indicating the disappearance of the after-sensation.

If this was not the case, the final signal appeared when there were no waves at all, or when small, fast ones were present.

TABLE VI

PERCENT OF TRIALS IN WHICH ALPHA WAVES WERE PRESENT BEFORE FINAL SIGNAL

Condition *A* (10 sec.)	68	$N=50$
Condition *B* (10 sec.)	79	$N=47$
Condition *A* (5 sec.)	82	$N=50$
Condition *B* (5 sec.)	84	$N=43$

Since our subjects did not report clean-cut disappearance and appearance of successive after-sensations but rather relatively gradual shifts from positive to negative and vice-versa, we were unable to check Jasper and Cruikshank's finding that brain potential activity is greater between successive after-sensations than during them (**3**). Our analyses revealed no consistent relation between changes in the alpha rhythm and changes in the after-sensations, except as we have just noted, namely, that in 90 percent of the trials signalling of the changes occurred during the absence of the rhythm. In a sense this is the opposite of what Jasper and Cruikshank found, but different experimental conditions and methods of analysis used in the two studies make comparison of findings of doubtful value.

During an after-sensation period, the mean number of bursts of alpha waves was 6.3 and the mean number of after-sensation changes 5.7. These values are quite similar but the bursts have no chronological relation to the changes. For example, at times there would be no alpha waves, while at other times there would be several bursts between two signals.

Frequency of Waves.—We chose subjects who had a dominant alpha rhythm. Three of them had a mean frequency of 9 and the other of 10 alpha waves per sec. No frequency differences were found between the alpha waves occurring before the light stimulus and those occurring during the after-sensation periods.

Summary and Conclusions

The effect of visual after-sensations upon alpha brain potentials was determined under two conditions, a high degree and a low degree of attentiveness to after-sensation changes. For the high degree of attention the total duration of alpha waves and the mean duration of the bursts during the after-sensation periods were less and the length of time for the first burst to appear after the light went off was greater than for the low degree of attention. These findings indicated a greater effectiveness in altering brain potential patterning of conditions of greater attention. It was calculated that the attentional factor contributed at least 35 percent of the total effect, the functional properties of the retina probably contributing the larger part of the remainder.

It is felt that certain psychological factors such as attention may be revealed in brain potential patterning and thus be objectively determined.

(Manuscript received December 20, 1937)

References

1. Travis, L. E., and Knott, J. R., Brain potential studies of perseveration: I. Perseveration time to light, *J. Psychol.*, 1936, 3, 97–100.
2. —— ——, Brain potential studies of perseveration: II. Perseveration time to visually presented words, *J. Exper. Psychol.*, 1937, 21, 353–358.
3. Jasper, H. H., and Cruikshank, R. M., Electroencephalography: II. Visual stimulation and the after-image as affecting the occipital alpha rhythm, *J. Gen. Psychol.*, 1937, 17, 29–48.

THE EFFECT OF SUBCUTANEOUS INJECTIONS OF BENZEDRINE SULFATE ON THE ACTIVITY OF WHITE RATS

BY LLOYD V. SEARLE AND CLARENCE W. BROWN

University of California

Introduction

The early studies of benzedrine were concerned primarily with its peripheral vascular effects. In connection with attempts to synthesize ephedrine, which had proved important in the treatment of respiratory ailments, several other related compounds, including benzedrine, were synthesized and studied. During one of these investigations Alles (**1**) demonstrated that benzedrine had a marked effect upon animals anesthetized with sodium amytal. The time of narcosis was greatly reduced by an administration of benzedrine. The recovery from the anesthetic appeared to result from a direct action of the drug on the central nervous system and not as an indirect result of the improved circulation, *i.e.*, increased blood pressure, which is a characteristic effect of benzedrine.

This finding led other workers to investigate the use of the drug in the treatment of persons suffering from narcolepsy and related conditions. Prinzmetal and Bloomberg (**4**) and Ulrich, Trapp, and Vidgoff (**8**) studied the effects of benzedrine on narcoleptic patients. Most cases experienced complete relief from attacks of sleepiness. Those who had been chronic narcoleptics immediately became wide-awake and energetic and remained so for several hours. Myerson, Loman, and Dameshek (**2**), working with mental disease patients at the Boston State Hospital, have extended the field of usefulness of the drug to include the treatment of the lowered mood and fatigue states accompanying certain psychoneurotic conditions, relaxing spasms of the gastro-intestinal tract, and counteracting an expected fall in blood pressure usually found

in certain types of anesthesia. Peoples and Guttman (3), working with both depressive and normal subjects at the Maudsley Hospital in London, report that marked psychic symptoms were exhibited following the administration of benzedrine. Normal subjects showed an increased motor restlessness and reported an experience of a change in mood, generally in the direction of euphoria. When a continuous addition test and a test of tapping rate were given following the injection of the drug, there was an increase in performance. In one study on these same subjects in which intelligence tests were used the group having benzedrine showed an average increase of seven points of I.Q. over the control group. These same investigators are conducting studies to test a theory of E. Krapf that the lability of the blood pressure is an important factor in the pathogenesis of psychoses, and they are finding benzedrine useful in producing systematic changes in the blood pressure conditions of their subjects.

It is evident from these studies that benzedrine is of considerable psychological interest, in view of the emotional, motivational and cognitive changes found to accompany its use. Its clinical value has been amply demonstrated. On the theoretical side there is some indication that it may offer a means of investigating the relation between blood pressure and emotional changes. It occurred to the present writers that the apparent effect of benzedrine on the central nervous system merited further investigation. This could best be studied in animals and the white rat was selected as the most suitable subject. Before beginning a direct attack on the problem it was thought that a more thorough knowledge of the general behavioral effects of the drug should be obtained. The present study was organized to determine the nature of the effect of the drug on the activity of the white rat. The revolving drum method of recording the behavior of the rat was utilized because it offered a rather accurate and reliable measurement of the animals' running activity.

Methods and Procedure

The experimental animals were fifteen white rats selected from four litters. Because of the fluctuations in activity incident to the oestrus cycle in females, only

males were used. Their ages at the beginning of the experiment ranged from 103 to 116 days. Since the experiment ran for a period of 61 days, all animals were tested between the ages of 100 and 200 days of age, which according to Shirley (7) is the time of highest spontaneous activity, the peak occurring at about 270 days. Although Shirley's results were based upon a larger number of rats, but under conditions more nearly similar to ours, our data are in closer agreement with Richter's (5) findings which showed a gradual decline in activity after the age of 100 days.

The revolving cages were made of ¼" 3-ply wood with the running surface of ⅜" mesh galvanized iron wire. Each cage was 8" wide and 14" in diameter, with a door on one side allowing access to the interior. They were mounted in blocks of five, three tiers high, and placed against a white wall. A white curtain was hung across the front, about four feet from the apparatus, to prevent undue interference from movement of persons in the room. Each cage was fitted with a Veeder counter which recorded revolutions in both directions and was so adjusted that approximately half a revolution had to be made in order to count a turn. This prevented inaccurate recording from frequent reversals, but probably was not an important precaution, since the rats, after getting started, showed a fairly consistent tendency to run in one direction. Water was always available to the animals. The rats were fed once a day, at 5:00 P.M., with the standard laboratory diet, which was mixed with water to such a consistency that it could be rolled in a ball of ample size and placed in the cage. Forty-five minutes were allowed for eating after which all uneaten food was removed. With the diet used the rats increased in weight during the experimental period from an average of 181 grams to an average of 250 grams.

Pure benzyl methyl carbinamine sulfate, dissolved in physiological salt solution, was given in subcutaneous injections. The strength of the solution was .3 mg per c.c.[1] Each animal received one c.c. of this preparation per 100 grams weight during the first half of the experiment, after which, because increasing weight necessitated the injection of very large amounts, the strength was doubled and only ½ c.c. per 100 grams was administered.

Because of the difficulty of equating groups in such a variable function as spontaneous activity, no control group was used. Instead, the drug was administered only once every four days, and since the physiological effects had been found to be dissipated within twenty-four hours (4), the intervening days were used as controls and the activity of these days compared with the activity of the drug day. As a control for variable factors of technique in handling the rats and the exciting effect of the injection, plain physiological salt solution was injected on one day of each 4-day period, the actual sequence of days being drug, rest, salt, rest, etc. In order that the activity incident to hunger would not interfere with the effects of benzedrine, all injections were made at 9:00 A.M., eight hours before feeding time. The counters were then read at one-hour intervals between 9:00 A.M. and 5:00 P.M.

The animals were allowed 21 days in which to become accustomed to the cages. This is far more than necessary for most problems and Shirley (7) considers ten to

[1] Since no studies had been reported in which benzedrine was used on rats, it was necessary to determine the dose experimentally. Subcutaneous and intraperitoneal injections were made on 17 rats, with doses ranging from 1.5 to 90 mg/kg. About 30 mg appeared to be the minimal lethal dose. A dose of 3 mg/kg was well below the lethal amount and yet produced marked behavioral changes in the animals. Results of this experiment will be made available to anyone interested.

fifteen days sufficient. In view of our method of control, it was desirable that the rats show a reasonably constant daily level of activity so that comparisons could be made within the 4-day periods. No such constancy was apparent in the raw data. However, a split-half reliability coefficient, computed for the ten-day period from the 9th to the 18th day, was .97 (corrected, .99) with a standard error of .015. This indicated consistent individual differences over the given time, although wide daily fluctuations continued throughout the entire experimental period. Beginning with the 11th day, five injections of salt solution were given on alternate days in order to accustom the rats to the injection procedure. This was followed by the sequence previously described, for a series of ten 4-day periods.

Results with Discussion

1. *Effect of 3 mg/kg Dose on Activity*

An examination of the hourly activity of the rats on drug days showed that the effects of the benzedrine had almost wholly disappeared after from four to five hours, or at about 1:30 P.M. Furthermore, the characteristic increase in activity preceding the time of feeding was noted to begin usually between 1:00 P.M. and 2:00 P.M. As this pre-feeding activity amounts to about twenty-five per cent of the total daily activity on rest days it was thought best not to allow it to dilute the activity incident to the injection of benzedrine. Comparions of the activity on the different days of each 4-day period are therefore stated in terms of the activity of the first 4 hours following the injection time or from 9 A.M. to 1 P.M. Table I gives the average daily activity for the forty days of the experiment. The condition of each day is indicated in column two. Column three gives the total 24-hour activity and column four the activity of the 4-hour period from 9 A.M. to 1 P.M.

It will be seen that benzedrine produces a very marked increase in activity in each of the 4-day periods. An increase also occurs on the salt days, indicating that the handling and injecting of the rats has a stimulating effect. Some idea of the magnitude of the drug effect, over and above the increase due to the procedure of injecting, can be gained from the standard critical ratios found in column five. Each ratio indicates the statistical reliability of the difference between the activity on the drug day and the salt day for a given 4-day period. All of the ratios are very large. For the ten 4-day

TABLE I

DAILY ACTIVITY

Day	Condition	24-hour	4-hour 9:00 A.M. to 1:00 P.M.	Critical Ratios of Differences between Drug and Salt Days
1	Salt	52,351	4,351	
2	Rest	58,576	1,960	
3	Drug	57,877	26,492	9.48
4	Rest	56,464	1,019	
5	Salt	39,286	3,933	
6	Rest	60,755	1,058	
7	Drug	56,200	33,159	7.46
8	Rest	53,483	884	
9	Salt	32,557	4,322	
10	Rest	65,641	362	
11	Drug	47,868	23,172	7.84
12	Rest	38,264	416	
13	Salt	50,258	3,710	
14	Rest	55,352	111	
15	Drug	42,062	22,002	11.41
16	Rest	31,014	27	
17	Salt	35,407	3,837	
18	Rest	31,002	422	
19	Drug	43,916	20,526	9.75
20	Rest	34,624	220	
21	Salt	47,312	2,297	
22	Rest	42,306	126	
23	Drug	37,792	18,670	8.47
24	Rest	46,844	1,340	
25	Salt	23,993	1,910	
26	Rest	27,090	151	
27	Drug	40,929	21,145	8.46
28	Rest	44,014	143	
29	Salt	44,347	3,032	
30	Rest	46,303	37	
31	Drug	33,160	17,625	6.13
32	Rest	27,281	100	
33	Salt	25,482	2,111	
34	Rest	54,827	1,308	
35	Drug	37,893	18,869	7.41
36	Rest	30,421	1,052	
37	Salt	27,654	2,062	
38	Rest	38,284	1,996	
39	Drug	36,939	15,991	8.20
40	Rest	39,451	1,212	

periods they range from 6.13 to 11.41 with a mean of 8.46. If the activity of the drug days is compared with the activity of the rest days the standard critical ratios are found to be of comparable size, ranging from 5.87 to 11.85 with a mean of 8.39.

The average 4-hour activity for all the salt days amounts to 210 revolutions. A similar average for the rest days is 26.5 revolutions. The activity of the salt days is then 7.9 times that found on the rest days. In the last two 4-day periods the amount of increase due to the salt injection is far less than for any of the other periods. This might be taken as an indication that the animals were becoming adapted to the injecting procedure and that eventually no marked increase in activity would have occurred from the injection of salt solution.

The average 4-hour activity on the drug days is 1451 revolutions per rat, or 54.8 times the activity of the rest days. The difference between 7.9 and 54.8, or 46.9, indicates roughly the relative amount of increase in activity which results from the administration of a 3 mg/kg dose of benzedrine. It should be stated that this does not indicate the increase in activity to be expected from the administration of an optimal dose of the drug to each animal, in view of individual differences in susceptibility to the drug. Further investigation of this topic will be reported in a later paper.

2. *Hourly Activity*

The average hourly activity from 9:00 A.M. to 5:00 P.M. is presented in Fig. 1. The figure is a composite of all ten 4-day periods, the rest, salt, and drug days being graphed separately. The maximum effect from the injection of benzedrine occurs during the first hour, although the averaging of the activity of all the rats tends to cover up large individual differences in the time of onset of the effect of the drug. After about four to five hours the effect of the drug wears off to the point that it cannot be separated from the increase in activity incident to hunger. The increase in activity resulting

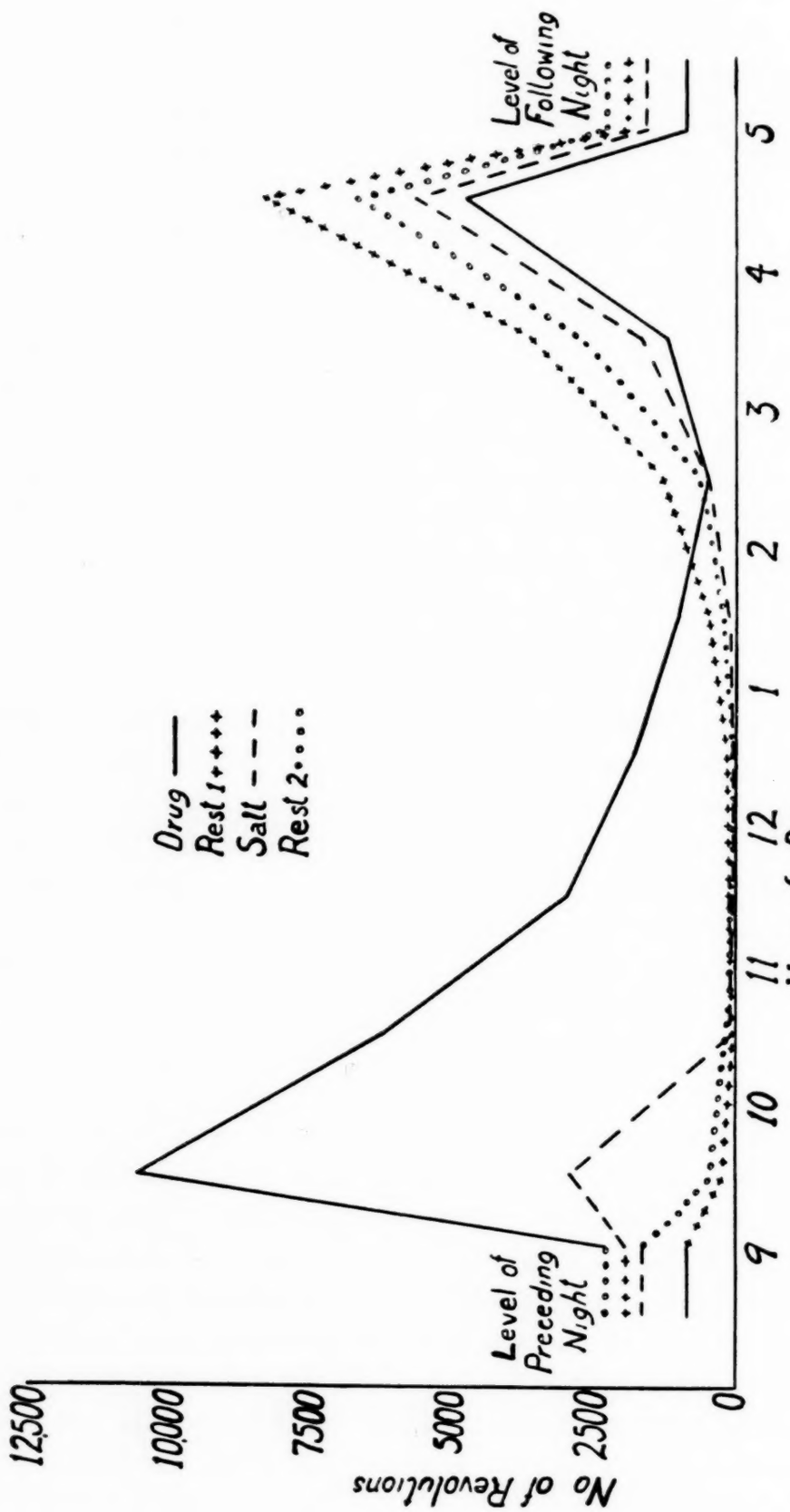


FIG. 1. Composite hourly activity curve.

from the injection of salt solution is seen to be negligible after the first hour.

The pre-feeding activity begins to show about 1:30 P.M. and from this time until the rats are fed there is a continuous increase in activity. It will be noted that the amount of the pre-feeding activity is closely related with the condition of that day and the preceding night. When the rats are made increasingly active due to benzedrine they show less pre-feeding activity. When, following a drug day, they become relatively inactive during the night and the early hours of the next day, they then show the greatest pre-feeding activity on the first rest day. The salt and second rest days show a similar sequence. It would appear from these results that the increased activity from benzedrine causes a reduction of activity at other times of the day. Normally the rats are more active at night, sleeping a good portion of the daytime. The daytime activity produced by the drug then causes a reduction in activity at night.

That the increase in activity from benzedrine results from a redistribution of the energy expression during a day's run is further substantiated by comparing the 24-hour activity record given in column three of Table I with the 4-hour record given in column four. If the increased activity due to benzedrine produced an absolute increase in activity for the day, then the 24-hour record should show an increase on each drug day. Likewise, there should be an absence of high peaks associated with rest days. It will be observed that the first, second, fifth, seventh, and tenth drug days are associated with relatively high levels of activity for the 24-hour period but that drug days three, four, six, eight, and nine are associated with relatively low levels of activity. Furthermore, a comparison of the 24-hour activity curve during the experimental period with the 24-hour curve during the preliminary adaptation period of 21 days shows that the same kinds of rises and falls occur in each. Apparently then benzedrine does not produce an increase in the absolute activity of the animals but compels them to redistribute their energy during each day.

3. *Tolerance for the Drug*

An examination of the activity on drug days shows that the amount of increase in the running becomes less as the experiment progresses; the number of revolutions for the 4-hour periods of the last three drug days is 37 percent less than that of the first three. Immediately one thinks of the animals developing a tolerance for the drug with the continued injections and offers this as an explanation for the decrease in the drug's effect. However, the factor of a general reduction in activity with increasing age must be taken into account before this conclusion can be considered valid. With increasing age rats under normal conditions become less active, the number of revolutions of the activity cages gradually diminishing (5). Apparently this phenomenon is present in the benzedrine experiment, since the total 24-hour activity level shows a decrease of 29 percent during the forty days. It will be noted that between this decrease and the decrease for the drug days there is a difference of 8 percent which might be taken as evidence for tolerance. However, since the drug period represents a higher level of activity than the average activity level it is reasonable to expect that the age increase would show relatively greater influence upon it than upon general activity. That this is actually the case seems probable in view of the fact that pre-feeding activity on rest days, another period of high activity level, shows a drop of 46 percent during the experimental period. A further evidence against the development of tolerance for benzedrine in the rat can be obtained from a consideration of the standard critical ratios given in Table I. If the benzedrine was having a relatively smaller effect on activity as the experiment progressed there should be a reduction in the reliability of the differences between the drug days and the salt or rest days. There is considerable variation in the critical ratios but they do not tend to get less as the experiment continues. The ratios for the last 4-day period, although not as large as those for the first, are as large as those for the second and third periods.

Since it has been shown that the effect of benzedrine is to produce a redistribution of the energy expression during a daily period rather than to affect significantly the 24-hour total, tolerance would appear as a consistent decrease in the ratio of the 4-hour period to the 24-hour period for drug days. Table II shows these per cents for all ten drug days. It will

TABLE II

DRUG ACTIVITY FROM 9:00 A.M. TO 1:00 P.M. EXPRESSED AS PERCENTS OF TOTAL 24-HOUR ACTIVITY ON DRUG DAYS

Period	4-hour 9:00 A.M. to 1:00 P.M.	24-hour	Percent	Averages of First and Last Three Percents
1	26,492	57,877	46%	
2	33,159	56,200	59	51%
3	23,172	47,868	48	
4	22,002	42,062	52	
5	20,526	43,916	47	
6	18,670	37,792	49	
7	21,145	40,929	52	
8	17,625	33,160	53	
9	18,869	37,893	50	49
10	15,991	36,939	43	

be noted that nearly 50 percent of the daily total activity occurs during the first four hours following injection of the drug, and there is no consistent trend towards reduction of this percentage during the experiment; the average of the last three days is only 2 percent less than that of the first three, a change not large enough to be considered significant.

CONCLUSIONS

1. A 3 mg/kg dose of benzedrine sulfate produced a marked rise in the spontaneous activity of white rats which endured for a period of approximately four hours following injection.

2. When salt solution was injected as a control for the exciting effect of the injection procedure, a slight rise in activity resulted during the first hour.

3. The increase in activity following injection of benzedrine is greatest during the first hour and falls off rapidly to normal in about four hours.

4. The pre-feeding and night activity levels following the 4-hour activity increase from benzedrine are consistently

lower than normal. Whether this is due to a depressant effect of the drug or to fatigue following over-exertion is not shown by the data.

5. No reliable indications for the development of tolerance for the drug during the experimental period of forty days was found.

Summary

An experiment was conducted to determine the effect of benzedrine sulfate on the spontaneous activity of white rats. Subcutaneous injections of a 3 mg/kg dose produced a marked increase in activity which lasted for approximately four hours and which was followed by a compensating decrease; the 24-hour drug day was not significantly different from salt and rest days. No tolerance was demonstrated for ten injections of drug over the experimental period of forty days.

(Manuscript received January 3, 1938)

Bibliography

1. Alles, G. A., The comparative physiological actions of the dl-β-phenylisopropylamines; I Pressor effect, *J. Pharmacol. Exp. Ther.*, 1933, **47**, 339–354.
2. Myerson, A., Loman, J., and Dameshek, W., Physiologic effects of benzedrine and its relationship to other drugs affecting the autonomic nervous system, *Am. J. Med. Sci.*, 1936, **192**, 560–674.
3. Peoples, S. A., and Guttman, E., Hypertension produced with benzedrine, *Lancet*, 1936, **1**, 1107–1109.
4. Prinzmetal, M., and Bloomberg, W., The use of benzedrine for the treatment of narcolepsy, *J. A. M. A.*, 1935, **105**, 2051–2054.
5. Richter, C. P., A behavioristic study of the activity of the rat, *Comp. Psychol. Monog.*, 1922, **1**, 2.
6. Searle, L. V., and Brown, C. W., The effect of injections of benzedrine on the activity of white rats, *Psychol. Bull.*, 1937, **34**, 558–559.
7. Shirley, M., Studies in activity; II. Activity rhythms; age and activity; activity after rest, *J. Comp. Psychol.*, 1928, **8**, 159–186.
8. Ulrich, H., Trapp, C. E., and Vidgoff, B., The treatment of narcolepsy with benzedrine sulphate, *Annals Inter. Med.*, 1936, **9**, 1213–1221.

A MATHEMATICAL ANALYSIS OF THE EXPERIMENTS IN EXTRA-SENSORY PERCEPTION

BY DONALD L. HERR

Moore School of Electrical Engineering, University of Pennsylvania

There has recently appeared a great amount of literature concerning the experiments in extra-sensory perception, in which the subject has been dealt with from a great many viewpoints. The purpose of this article is to attempt to clarify some existing problems, solve some of them, and to make suggestions for further investigation.

No matter what one may think of the conclusions drawn from Dr. Rhine's work on telepathy and clairvoyance, his statistical bases must be beyond reproach before his results can be properly weighed. Accordingly, then, the first problem to be discussed is the nature of the statistical 'universe' which is encountered in the Rhine experiments.

The composition of Dr. Rhine's deck is fully known to the subject, therefore, his calls tend to fall more or less in the same suit distribution, namely five 'calls' of each suit. This tendency much restricts the subject's freedom in calling, whether he recognizes it or not. More important, however, this tendency makes it impossible for the hits (coincidence of call and card turned up) to follow the binomial distribution, and hence no arguments or calculations based on the binomial distribution are justified.

An analogy for a four-card, two-suit deck is appended to this paper. This deck was chosen for study because it is small enough for all possible permutations to be considered in a reasonable lifetime.[1] It is recognized that it is a special example, and hence does not fully demonstrate the exact distribution followed by Dr. Rhine's deck, although it does give a good hint of its nature. The appended example shows

[1] These two sections are the result of joint work with Dr. H. F. Smyth, Jr., of the University of Pennsylvania, to whom most of the credit belongs for them.

that when the subject adheres to calls of the same composition as the deck (two cards in each of two suits) he will hit all four cards once in 6 trials, not once in 16 as the binomial distribution demands. The total number of hits in 100 trials is, however, 50, the number calculated from the probability per call of 1/2. This is due to the fact that zero hits are correspondingly more frequent than the binomial distribution allows.

For analogous reasons, no calculation of 'p,' the chance for success, based on the total number of permutations possible and on the binomial expectancy of calls matching cards turning up is justified.

Any computation of σ, the standard deviation, taken equal to $\sqrt{pqn}$ is also unjustified because the formula is only true for a binomial distribution of 'hits.'

Dr. Rhine's statement that the probability of a 'hit' on each 'call' is 1/5 can easily be shown mathematically, although some have doubted it. There now follows the discussion of the probability of guessing any card right:

Card matching in successive shuffles of such a pack need not be repeated many times to convince any experimenter that the probable number of hits is 5 per 25 trials. This is exactly analogous to the subject calling exactly five cards of a suit, with all possibility of telepathy and clairvoyance removed. Dr. Rhine reports that he tried this 4,000 times, and found 801 hits, where he had expected 800. I have tried it only 49 times, with 241 hits in 1,225 comparisons, or an average of 4.93 per 25 comparisons. Although probable error of this mean cannot be calculated due to the departure from binomial distribution of such restricted matching, yet the hits are close enough to a figure of 5 per pack of 25 to be quite convincing.

I insert here a rigid mathematical proof that the probability of a hit is exactly the same on every card of the deck.

Given a pack of N cards, divided into P suits of Q cards in each (Q may or may not equal P). The pack is shuffled, and the subject calls the suits with no knowledge of what suits are turned up, but it is conceded that he may have full knowledge of the composition of the pack; that is, he may know that $N = PQ$. It is of course immaterial whether he calls the suits one by one as the pack is leafed through, or if he calls them 'down through' without the cards being handled. The suit called for the first card of the pack will not be determined, and in the long run will be equally distributed among

the P suits. Although each successive call may be determined by the first call in the subject's efforts to keep his calls evenly divided among the suits, yet since the first call is not determined in the long run, no subsequent call will be determined either, no matter what 'system' the subject follows. It follows that there will be an equal chance, $1/P$, of his calling any one of the P suits on each call.

In the following tabulation, $r, s, \cdots$ represent the number of cards remaining in the deck just before the nth call, in the respective suits $A, B, \cdots$. $r, s, \cdots$ may be any integers from O to Q, and obviously their sum must be $(N - n + 1)$ since their sum is the number of cards remaining in the deck just before the nth call.

Tabulation: For the nth call

Suit	Cards Yet to Turn Up	Times Suit Is Called on the nth Call	Times Suit Turns Up	Times Suit-Call Is a 'Hit'
A	r	$1/P$	$r/(N-n+1)$	$r/P(N-n+1)$
B	s	$1/P$	$s/(N-n+1)$	$s/P(N-n+1)$
etc				
Total for P suits	$r+s\cdots$	P/P	$r+s\cdots/(N-n+1)$	$r+s\cdots/P(N-n+1)$
or	$(N-n+1)$	1	$(N-n+1)/(N-n+1)$	$(N-n+1)/P(N-n+1)$
or			1	$1/P$

Thus, it is seen that the probability of success on the nth call is $1/P$ (it makes no difference if Dr. Rhine's deck of five suits of five cards is used, or another deck of say, five suits of ten cards each). Therefore, in Dr. Rhine's deck, the chance of success on each and every call is 1/5.

The discussion now evolves about this point: Is there some criterion which may be applied, once the standard deviation of the frequency distribution of hits, with chance alone acting, is known? Fortunately, there is, although it appears that obtaining the standard deviation of the frequency distribution of hits, with *ESP* not operating, and with the subject knowing the form of the deck beforehand, is about as difficult as determining the exact distribution itself. Empirically the process would be impractical. There has appeared, however, an article by Sterne [2] in which he gives exactly the 2d, 3d, 4th moments about the mean of 5. The distribution is that of Pearson's Type I. Using these moments, as developed by Dr. Sterne, I have computed, according to Pearson's method,[3, 4] the standard deviation of this distribution. The value for σ is

[2] T. E. Sterne, *Science*, 1937, **86**, 500.

[3] Karl Pearson, Mathematical Theory of Evolution, Philos. Trans., **186**, Part 1, 343 to 414.

[4] T. C. Fry, *Probability and Its Engineering Uses*. Appendix.

2.041 the last figure being accurate to within 1 or 2 digits.[5] This is the same as $\sqrt{\mu_2} = \left(\frac{25}{6}\right)^{\frac{1}{2}}$, μ_2 being the value of the second moment as given by Sterne. We are now in a position to put Dr. Rhine's data to an exact test to determine whether or not *ESP* exists.

Tschebyscheff developed an inequality which bears his name, which, applied *to any distribution whatsoever*, states that P_z, the proportion of the total number of events having a value X (measured from the mean, in this case 5) less than $Z\sigma$, where Z is some number greater than 1, and σ is standard deviation, is greater than $1 - \frac{1}{Z^2}$; or, $P_z > 1 - \frac{1}{Z^2}$. If we choose $Z = 3$, $P_z > 0.89$; that is, *at least* 89% of the observed trials would be in a range 3σ about the mean. P_z, for any Z, is now readily obtainable from Sterne's Pearson Type I distribution fit, and therefore the criterion may be applied for any range $Z\sigma$ about the mean. If more than 11 percent of Dr. Rhine's results exhibited values greater than 3σ, σ being computed from the distribution encountered on a pure chance basis and here equal to 2.041, then one of two deductions is possible. Either *ESP* exists, or else some of the mediocre, normal results have not been included by the experimenters in the course of the study. If the latter has been done, the results to date are without value whatsoever; at the present no detailed report on this matter, nor a detailed account of the collected statistics, has appeared in print to the author's knowledge.

Getting back to the point considered, if the frequency distribution of Rhine's total accumulated data lies within or is congruent with Sterne's chance distribution fit, then *ESP* may be said not to exist. From this discussion, it is seen that the mechanics of the problem afford a somewhat awkward mathematical problem. *ESP*, if it exists, should not be a delicate function of method. A wise solution would be a

[5] Dr. Harold Lufkin, of the Mathematics Department of the University of Pennsylvania, and Dr. C. Winston, of the National Resources Committee, have kindly assisted in the checking of this work.

continuation of the *ESP* experiments under a method whose mathematical analysis is somewhat simpler. The following two alternatives are suggested:

(1) Have 25 identical sub-decks, each of 5 cards, 1 of each suit. Shuffle each sub-deck separately, then take the top card of each sub-deck so as to construct the major deck of 25 cards. Any distribution of the 5 suits is now possible, some suits can be lacking, or there may appear a distribution of the nature Dr. Rhine now employs. Never need the subject subconsciously limit his calls to 5 in each suit. The 'hit' distribution will now most certainly be binomial. Hence all the well-known binomial statistics may be employed.

(2) Another arrangement permitting the binomial analysis follows: Have a deck of 5 cards, one of each suit, shuffle this each time, and have the subject name only the top card after each shuffle.

Of course, some objection may be brought forth questioning the necessity for a change of method, especially after some 250,000 trials have been undertaken by some other method. *ESP*, if it exists, should reveal itself under the application of a variety of methods, and, indeed, the methods mentioned above yield results more readily comprehended by the plainer mind than those heretofore brought forth.

If Dr. Rhine's experiments seem to establish that something new and so fundamental has been revealed, certainly further confirmation of its existence by additional tests can be the only next step, if our processes are to be truly scientific.

(Manuscript received February 24, 1938)

APPENDIX

Demonstration in a simple example of departure from binomial distribution of 'hits' in guessing a four card two suit deck.

Deck: four cards of two suits, A, A', B, B'

These cards can be arranged in ${}_4P_4$ ways $= 24$

Considering suits only, they can be arranged in ${}_4P_{4(2,\,2)}$ ways $= 6$

For each shuffle (arrangement of cards) the subject guesses the suit arrangement. In the tabulation below the subject's 'hits' are marked with an 'x.'

Permutations of Deck		All Possible Calls of Subject Appear at Tops of Columns															
		a. Subject adheres to known composition of deck.						*b*. Additional calls possible when subject pays no attention to composition of deck									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		A	A	A	B	B	B	A	A	A	B	A	B	B	B	B	A
		A	B	B	A	A	B	A	A	B	A	A	B	B	B	A	B
		B	A	B	B	A	A	A	B	A	A	A	B	B	A	B	B
		B	B	A	A	B	A	B	A	A	A	A	B	A	B	B	B
1′	*A*	x	x	x				x	x	x		x					x
	A′	x			x	x		x	x		x	x				x	
	B	x		x	x				x				x	x		x	x
	B′	x	x			x		x					x		x	x	x
2′	*A*	x	x	x				x	x	x		x					x
	A′	x			x	x		x	x		x	x				x	
	B′	x		x	x				x				x	x		x	x
	B	x	x			x		x					x		x	x	x
3′	*B*				x	x	x				x		x	x	x	x	
	A	x			x	x		x	x		x	x				x	
	B′	x		x	x				x				x	x		x	x
	A′			x	x		x		x	x	x	x		x			

The other 21 possible permutations of the deck obviously yield results identical in kind.

Notice that when the subject adheres to the known composition of the deck (*a*) he makes 4 hits in one set of four guesses once out of six sets, 0 hits once out of six sets, and 2 hits four times out of six, and he cannot make 1 hit or 3 hits. On the other hand, when he guesses freely without attention to the composition of the deck (*a* and *b* together) he makes 0 hits once out of 16 sets, 1 hit four times, 2 hits six times, 3 hits four times, and 4 hits 1 time. The latter is the distribution of hits demanded by the binomial law; the former departs widely from the binomial distribution. However, with each type of guessing, the subject makes exactly one hit out of two guesses.

This results from the fact that binomial distribution can be followed only when the events considered are completely independent, as they certainly are not when a subject adheres to any set ratio between guesses of different suits.